The Effect of Chairman’s Statement Tone Changes in Annual Reports from Hong Kong

Junjie Li\textsuperscript{1}, Wang Luo\textsuperscript{2} and Xin Deng\textsuperscript{3}
\textsuperscript{123}College of Computer Science and Software Engineering, Shenzhen University, Guangdong China

Abstract. The tone captures how the leaders of listed company confidence to the performance and tone changes correlate with revisions of future outlook. Current predicting stock market behaviour is using numerous quantitative financial factors. Recent publications have demonstrated that some implied sentiment information such as tone changes in annual reports can be successfully used to predict the stock price in the U.S market. However, the investors' reflection to the tone changes in annual reports in Asia market, especially in Asian financial center Hong Kong, is still unknown. In this paper, the chairman's statement tone changes in annual reports from the Hong Kong market have been studied in the first time. This study evaluates three different tone changes methods and combing with financial indicators to predict the stock price. The experimental results prove that the tone changes of annual reports can predict the stock price in the long trend, which implies the low market efficiency in Hong Kong. Moreover, some experiments have been investigated whether the financial crisis can be predicted from the chairman's tone changes.

1. Introduction
Predicting stock market behaviour is attracting numerous attentions [1]. Some attempts concentrated on discovering the causal relationships from quantitative factors, such as BM, Market Capitalization, Turnover and so on. Eugene F.Fama got the Nobel prize in Economic for his famous Fama-French three-factor model [2].

Stock price reactions not only correlate with the quantitative measures of firm fundamentals. Sentiment analysis for the financial document, such as newspaper, has increased dramatically in behavioural finance in the past decades while the availability of online newspaper and electronic financial documents. Studies have shown the sentiment in financial documents has a strong correlation with the future stock performance [3,4,5].

The annual reports disclosure the activities and financial performance of the listed companies throughout the preceding financial year. Several attempts analyze sentiment information in annual reports and forecast stock prices for the U.S market. Loughran and Mcdonald [6] observe that the stock price will move down around the filing dates if the firms convey more negative tone in their 10-K filings. T Arnold [7] quantifies the words used in the risk factors sections of IPO prospectuses and find that they capture most of the IPO returns. Yan [8] shows that the tone of 10-Ks gives accurate predictions of firms’ long-term performance. Yousry Ahmed [9] finds that the tremendous influence of the textual sentiment of corporate disclosure for the forecasting of corporate investments and financing decisions.

Recently, some studies explore the information content and value relevance of the tone and sentiment conveyed through the annual report. Kanget al.[1] found positive report tone and earnings
persistence having a positive relationship in the U.S. market. Feldman et al. [10] indicates that the short window market reactions are significantly associated with the tone of the MD&A section in 10-Ks. Maik Schemeling et al. [20] shows that changes in the tone of financial news and communications can forecast the price movement. The above literatures show that words in textual corporate communications have significant predictive power for firms’ future stock returns, earnings, and performance. However, the investors’ reflection to the tone changes in annual reports in Asia market, especially in Asian financial center Hong Kong, is still unknown. This work will fill gaps for sentiment analysis of the Chairman’s Statement section in the annual report from the Hong Kong market.

This paper presents a novel sentiment analysis approach for the Asian market using tone changes measures and verify its validity. The work's main contributions are: i) It is the first work to analyze the sentiment of the financial annual reports in the Asian market. ii) It is the first work to study tone changes in the Chairman’s statement section of listed companies’ annual report. iii) Studying the implication of tone changes for the trend of stock price, shows that the combination of tones and financial quantitative data can better predict the long trend of stock price. iv) Finding characteristics of Hong Kong’s financial market by comparing the accuracy for tone changes to predict stock price trend in different cycle span.

The rest of the paper is organized as follows: The next section reviews the relevant literature and motivates. Section 3 describes the sample, defines and describes methods of computing tone changes used in the paper. Section 4 presents experiments and results and Section 5 concludes the paper.

2.Literature Review

In the financial domain, the investor information is classified into two categories, one category that is quantifiable, such as size, market capitalization, turnover and so on, and another category that is non-quantifiable, such as news, annual report, announcement and so on [10].

Recently, scholars pay most of attention to the study of financial texts. Das [3] studies small investor sentiment based on stock message boards, which is then used to assess the impact on investor opinion of management announcements, press releases, third-party news, and regulatory changes. You and Zhang [4] finds that the annual report contains useful information about future company performance and abnormal stock price and trading volume changes near the issue date are positively correlated with future company profitability. Zhang and Skiena [5] performs a comprehensive and comparative study on the relation between news variables and the company's stock trading volumes and financial returns based on blogs and news, and then gives a sentiment-based market-neutral trading strategy which gives consistently favourable returns with low volatility over a long period.

Researches on financial non-quantifiable information concentrate on sentiment analysis of text documents. At present, sentiment analysis methods can be divided into two types, the first method is using word categorization, and the second method is based on statistical method. The former method requires a high-quality dictionary, which not only can give their categorization according to the sentiment of words, but also is sensitive to context. Loughran and Mcdonald [6] builds a financial-specific dictionary based on a large sample of 10-Ks during 1994-2008. The dictionary uses 6 word classifications (positive, negative, uncertain, litigious, strong modal, and weak modal) to measure the financial text sentiment. The latter methods require the likelihood ratios to be estimated based on subjective classification of texts tone. Schumaker and Chen [11] compares three text processing representations combined with support vector machines to predict stock prices and find the method based on named entities the most robust. Hjek et al. [12] applies several neural networks and support vector regression models to sentiment words of annual reports to predict the yearly change in the stock price of US firms. Chen et al. [13] builds tagging models based on the conditional random field (CRF) techniques to identify opinion holders. Deng et al. [14] proposes a supervised term weighting scheme based on importance of a term in a document and importance of a term for expressing sentiment, which employs seven statistical functions to learn the importance of a term for expressing sentiment of each term from training documents with category labels.
As for the approaches of sentiment quantification method in financial annual report (sometimes called 10-K in US), there are mainly two formats: single value or sentiment vector. Jegadeesh and Wu [15] defines a tone score for each document, whose value is the sum of the product of each sentiment word group and its frequency. Hjek and Bohov [16] and Kang et al. [1] calculate the overall tone for each document, defined as the count of positive words minus the count of negative words, divided by the sum of both positive and negative word counts (or divided by the total words counts of the document). The method, which uses a single value to represent sentiment of a text, is simply computed but not enough to explain text emotions. So some scholars start to try sentiment vector. Chu et al. [17] calculates the estimated overall market-sentiment vector for the commentaries by considering sentiment-linked topics. Wang Y [18] generates a 37 dimension sentiment vector based on 6 word categories of Loughran and McDonald [6] and 31 additional categories of Diction 5.0 of R.P.Hart [19].

However, these traditional sentiment vectors don't consider the meaningless emotions introduced by writing habits.

Tone changes method makes up for the shortcomings of traditional sentiment vector by considering the differences in words between two documents. Feldman et al. [10] measure tone changes in a specific Management Discussion and Analysis (MD&A) section of 10-K relative to prior periodic SEC filings by a well-established classification scheme of words into positive and negative categories. Schmeling and Wagner [20] firstly compute the tone of each annual report by 1 minus the ratio of the number of negative words and then measure tone changes by difference of two consecutive annual reports’ tone in adjacent years.

3. Data and Research Methodology

The annual reports of the listed companies came from Hong Kong Stock Exchange. This paper focuses on stocks belonging to Mainland private enterprises (MPE) and overseas-funded enterprises stocks and Hong Kong enterprises (OEE). There are 186 listed companies selected, including 79 MPE and 107 OEE candidates, which are from different sectors, such as Basic Materials (BMA), Communications (COM), Consumer-Cyclical (COC), Consumer-No-cyclical (CNC), Diversified (DIV), Energy (ENE), Financial (FIN), Industrial (IND), Technology (TEC) and Utilities (UTI).

Table 1. Mainland private enterprises (MPE) and overseas-funded enterprises and Hong Kong enterprises (OEE) Annual report distribution (OEE/MPE)

|     | BMA  | COM  | COC  | CNC  | DIV  | ENE  | FIN  | IND  | TEC  | UTI | total |
|-----|------|------|------|------|------|------|------|------|------|-----|-------|
| 2003| 0/1  | 1/0  | 5/5  | 2/4  | 4/0  | 0/1  | 17/1 | 2/0  | 1/1  | 2/0 | 34/13 |
| 2004| 0/1  | 1/0  | 7/8  | 1/4  | 4/0  | 1/1  | 19/1 | 3/0  | 1/1  | 3/1 | 40/17 |
| 2005| 0/2  | 2/0  | 8/10 | 1/4  | 4/0  | 1/1  | 21/3 | 5/2  | 1/1  | 3/2 | 46/25 |
| 2006| 0/2  | 3/1  | 8/9  | 2/6  | 5/0  | 1/1  | 23/3 | 5/3  | 1/3  | 4/2 | 52/30 |
| 2007| 0/2  | 3/1  | 9/14 | 2/6  | 5/0  | 0/1  | 22/6 | 5/5  | 1/2  | 3/2 | 50/39 |
| 2008| 0/4  | 4/1  | 11/14| 2/7  | 5/0  | 0/2  | 24/9 | 5/6  | 1/2  | 4/2 | 56/47 |
| 2009| 0/4  | 4/2  | 9/14 | 2/9  | 4/0  | 0/2  | 22/12| 5/8  | 1/2  | 4/2 | 51/55 |
| 2010| 0/4  | 4/2  | 12/13| 2/10 | 4/0  | 0/1  | 24/14| 4/8  | 1/1  | 4/2 | 55/55 |
| 2011| 1/5  | 4/2  | 14/16| 3/10 | 4/0  | 0/4  | 23/19| 5/9  | 1/2  | 4/2 | 59/69 |
| 2012| 1/5  | 5/2  | 13/15| 3/14 | 5/0  | 0/4  | 28/20| 4/11 | 1/2  | 4/2 | 64/75 |
| 2013| 1/5  | 5/2  | 15/14| 4/14 | 5/0  | 0/3  | 29/23| 4/11 | 1/2  | 4/2 | 68/76 |
| 2014| 1/6  | 6/2  | 15/17| 4/14 | 5/0  | 0/2  | 29/25| 5/11 | 1/4  | 4/2 | 70/83 |
| 2015| 1/5  | 6/2  | 18/15| 4/15 | 5/0  | 0/5  | 29/28| 6/11 | 1/2  | 4/2 | 74/85 |
| 2016| 2/6  | 6/3  | 17/20| 4/17 | 5/0  | 1/4  | 30/28| 6/12 | 1/4  | 5/2 | 77/96 |
| total| 7/52 | 54/20| 161/184| 36/134| 64/0 | 4/32 | 340/192 | 64/97 | 14/29 | 52/25 | 796/765 |

The financial indicators of the listed companies in the annual reports is retrieved from the Bloomberg terminal, which have shown significant impacts on the tone of annual reports in the
literature [15, 21, 22]. In this paper, natural logarithm of the market capitalization of equity (Size), ratio of the book value of equity (BM), standard deviation of the firm-specific component of returns (Volatility), Natural logarithm of the number of shares traded (Turnover), net assets per share (BPS), Equity Multiplier (EM), return on equity diluted (ROE), price-earnings ratio (PE) will be the fundamental financial indicators, which are used to predict the future performance of stock price.

Abnormal return is used to measure the reaction of stock on short period to financial market sentiment, such as 1 day [23], 3 days [16, 21] and so on. To analysis the reaction period of tone for Hong Kong financial market, the paper adopted abnormal return for stock price on short period (3 days, 5 days, 10 days and 20 days) and the trend of stock price on long period (60 days, 120 days and 250 days). All stocks’ day line price data range from 2000 to 2018 is collected, which are all in candidate stocks set. According to the computing method of abnormal return of Hjek and Bohov [16] and Jegadeesh and Wu [15], the paper selects Close price, the stock price data comes from futunn website, and consider Close-to-Close price return, which is

\[
\text{ret}_{it} = \frac{\text{Close}_{it}}{\text{Close}_{i,t-1}} - 1
\]  

\[
R_i = \sum_{t=0}^{n} \text{ret}_{it} \sum_{t=0}^{n} \text{ret}_{hst, it}
\]

\[
I_i = \begin{cases} 
1 & \text{if } R_i \geq 0 \\
-1 & \text{if } R_i < 0
\end{cases}
\]

where \( i \) stands for the annual report \( i \), \( t \) stands for the release date of Annual report \( i \), \( n \) is days interval, \( \text{Close}_{it} \) is close price of annual report \( i \) in \( t \), \( \text{ret}_{it} \) is the return of the stock for annual report in \( t \), \( \text{ret}_{hst, it} \) is the return for HSI in \( t \), \( R_i \) is the abnormal return of stock for annual report \( i \). \( I_i \) is the label for annual report \( i \). The value of \( n \) is generally 3,5,10 or 20 trading days considering previous researcher’s experience. As for the trend of stock price for long period, the paper use regression analysis to the long period stock price trend. The label is 1 when the slope of the regression line is positive, otherwise -1. All stocks were categorized into two classes on each year, with positive (932) and negative (629), indicating no obvious imbalanced problem.

The Chairman’s Statement section of annual report is used to interpret the important leadership communication made to all shareholders [23], which provides a high-level overview of highlights of company’s performance in the year gone by together with a comment on the current and overall scenario. That section shows a clear vision of the company and a broad view of the industry and national economy of the Chairman and other Managing Directors. Considering the size of financial-specific dictionary and different inflections of a word may have different sentiment, the paper selects the word group developed by Jegadeesh and Wu [15] which group different inflections of a word together and avoid the process of lemmatization and stemming. The word groups list contains 718 negative word groups and 123 positive word groups.

Table 1 shows the distribution of the number of Annual reports. The work prepares the original data (pdf text) by firstly extracting Chairman’s statement section from all pdf, and then converting all words to lower case and removing invalid characters, numbers, punctuation and extra whitespace, and stop words. After preparing the text files, three most common methods are selected to calculate annual report tone changes (also called delta tone).

**Schmeling Method**, the paper firstly quote the method of Schmeling and Wagner [20] to compute tone. Instead of only consider negative words in Schmeling method, the paper separately uses positive words and negative words to compute two kinds of tone. Firstly, classifying words as negative or positive in financial contexts by word groups of Jegadeesh and Wu [15], and then count the number of sentiment words in each txt and compute the ratio of the number of sentiment. The equation is

\[
\tau_i = 1 - \frac{N_{i, \text{senti}}}{N_{i, \text{total}}}
\]

\[
\Delta \tau_i = \tau_i - \tau_{i-1}
\]
Table 2. The effect of tone on different trading days for Bernoulli Bayes on all data

| Trading days | Schmeling | Kang&Feldman | Tetlock |
|--------------|-----------|--------------|---------|
|              | ∆ ∆ +Fin  | ∆ ∆ +Fin     | ∆ ∆ +Fin|
| Panel A: delta tone computed by negative words | 3 | 0.5239 0.5088 0.4962 | 0.5113 0.5063 | 0.5088 0.4962 |
|              | 5 | 0.5416 0.5239 0.5214 | 0.5239 0.5315 | 0.5264 0.5290 |
|              | 10 | 0.5290 0.5315 0.5290 | 0.5264 0.5239 | 0.5315 0.5340 |
|              | 20 | **0.6121** | 0.5315 0.5340 | 0.5315 0.5315 |
|              | 60 | 0.5063 0.5139 0.5642 | 0.5139 0.5642 | 0.5139 0.5668 |
|              | 120 | 0.5264 0.5416 0.5819 | 0.5416 0.5793 | 0.5416 0.5819 |
|              | 250 | 0.5139 **0.6121** | **0.6423** | **0.6095** | **0.6499** | **0.6121** | **0.6423** |
| Panel B: delta tone computed by positive words | 3 | 0.5063 0.5088 0.4912 | 0.5113 0.5063 | 0.5088 0.4912 |
|              | 5 | 0.5264 0.5264 0.5189 | 0.5239 0.5315 | 0.5264 0.5164 |
|              | 10 | 0.5239 0.5239 0.5239 | 0.5264 0.5239 | 0.5239 0.5264 |
|              | 20 | 0.5290 0.5264 0.5567 | 0.5315 0.5416 | 0.5264 0.5542 |
|              | 60 | 0.5139 0.5139 0.5819 | 0.5139 0.5642 | 0.5139 0.5819 |
|              | 120 | 0.5416 0.5416 0.5844 | 0.5416 0.5793 | 0.5416 0.5844 |
|              | 250 | **0.6121** | **0.6121** | **0.6373** | **0.5995** | **0.6499** | **0.6121** | **0.6423** |

Table 3. The effect of tone on different trading days for Bernoulli Bayes on negative words

| Trading days | Schmeling | Kang&Feldman | Tetlock |
|--------------|-----------|--------------|---------|
|              | ∆ ∆ +Fin  | ∆ ∆ +Fin     | ∆ ∆ +Fin|
| Panel A: OEE | 3 | 0.5049 0.5049 0.5196 | 0.5049 0.5147 | 0.5049 0.5147 |
|              | 5 | 0.5147 0.5147 0.5490 | 0.5098 0.5637 | 0.5098 0.5539 |
|              | 10 | 0.5245 0.4951 0.4853 | 0.4853 0.4902 | 0.4951 0.4804 |
|              | 20 | 0.4657 0.4706 0.5098 | 0.4706 0.5049 | 0.4706 0.5098 |
|              | 60 | 0.5147 0.5196 0.5784 | 0.5147 0.5784 | 0.5196 0.5784 |
|              | 120 | 0.6029 0.5539 0.5931 | 0.5686 0.5931 | 0.5539 0.5931 |
|              | 250 | **0.6373** | **0.6422** | **0.6471** | **0.6324** | **0.6422** | **0.6422** |
| Panel B: MPE | 3 | 0.4922 0.5233 0.4508 | 0.4819 0.4922 | 0.5181 0.4508 |
|              | 5 | 0.4922 0.4922 0.4560 | 0.4870 0.4715 | 0.4974 0.4560 |
|              | 10 | 0.4870 0.4870 0.4974 | 0.4611 0.4922 | 0.4922 0.4974 |
|              | 20 | 0.5389 0.5389 0.5648 | 0.5389 0.5026 | 0.5389 0.5596 |
|              | 60 | 0.4560 0.4611 0.5440 | 0.4560 0.5337 | 0.4611 0.5492 |
|              | 120 | **0.5596** | **0.5596** | **0.5751** | **0.5699** | **0.5803** | **0.5440** | **0.5699** |
|              | 250 | 0.5440 0.5440 | **0.6321** | 0.5440 0.6114 | **0.5440** | **0.6321** |
In Eq.(4) and Eq.(5), \( N_{\text{senti}} \) is the number of sentiment words, which is positive words or negative words in experiment. \( N_{\text{ltotal}} \) is the total number of words of chairman’s statement section of annual report \( i \). \( \tau_i \) is tone of annual report \( i \). \( \Delta \tau_i \) is the tone change of annual report \( i \) based on annual report \( i-1 \).

**Kang&Feldman Method**, the paper also applies the computing tone method of Kang [1] and Feldman [24]. The equation is Eq.(6). Where \( i \) is the annual report \( i \), \( N_{\text{in}} \) is the number of negative words, \( N_{\text{ipos}} \) is the number of positive words. The method of calculating tone change is like Eq.(5).

\[
\tau_i = \frac{N_{\text{lang}}-N_{\text{ipos}}}{N_{\text{ltotal}}} \quad (6)
\]

**Tetlock Method**, the paper also uses the method of Tetlock et al. [25] to compute the tone of annual report. Instead of only consider negative words like Tetlock, the paper separately use positive words and negative words to compute two kinds of tone. The equation is Eq.(7) and Eq.(8). Where \( \mu_{\text{senti}} \) is the mean of \( T_i \), \( \sigma_{\text{senti}} \) is the standard deviation of \( T_i \). The method of calculating tone change is like Eq.(5).

\[
T_i = \frac{N_{\text{senti}}}{N_{\text{ltotal}}} \quad (7)
\]

\[
\tau_i = \frac{T_i-\mu_{\text{senti}}}{\sigma_{\text{senti}}} \quad (8)
\]

Table 4: The effect of tone on different model for 250 trading days on all data

| Model       | Fin | Schmeling | Kang&Feldman | Tetlock |
|-------------|-----|-----------|--------------|---------|
|             |     | \( \Delta \) | \( \Delta +\text{Fin} \) | \( \Delta \) | \( \Delta +\text{Fin} \) | \( \Delta \) | \( \Delta +\text{Fin} \) |
| **Panel A:** delta tone computed by negative words |     |          |              |         |          |              |         |
| BNB         | 0.6121 | 0.6121 | 0.6423 | 0.5995 | 0.6499 | 0.6121 | 0.6423 |
| DT          | 0.5995 | 0.5441 | 0.5894 | 0.5013 | 0.5995 | 0.5164 | 0.6071 |
| GNB         | 0.5995 | 0.5869 | 0.5995 | 0.5693 | 0.5945 | 0.5894 | 0.6020 |
| LR          | 0.6121 | 0.6121 | 0.6247 | 0.6096 | 0.6247 | 0.6121 | 0.6247 |
| SVC-linear  | 0.6121 | 0.6121 | 0.6222 | 0.6096 | 0.6146 | 0.6121 | 0.6222 |
| SVC-rbf     | 0.5995 | 0.5995 | 0.6196 | 0.6020 | 0.6196 | 0.5516 | 0.6196 |
| SVC-sigmoid | 0.5567 | 0.5567 | 0.4962 | 0.5214 | 0.4962 | 0.4937 | 0.4861 |
| **Panel B:** delta tone computed by positive words |     |          |              |         |          |              |         |
| BNB         | 0.6121 | 0.6121 | 0.6373 | 0.5995 | 0.6499 | 0.6121 | 0.6423 |
| DT          | 0.5995 | 0.5768 | 0.5945 | 0.5088 | 0.6020 | 0.5365 | 0.5970 |
| GNB         | 0.5995 | 0.5894 | 0.5945 | 0.5693 | 0.5945 | 0.5995 | 0.5970 |
| LR          | 0.6121 | 0.6121 | 0.6247 | 0.6096 | 0.6247 | 0.6121 | 0.6297 |
| SVC-linear  | 0.6121 | 0.6121 | 0.6096 | 0.6096 | 0.6146 | 0.6121 | 0.6171 |
| SVC-rbf     | 0.5995 | 0.6020 | 0.6196 | 0.6020 | 0.6196 | 0.5164 | 0.6071 |
| SVC-sigmoid | 0.5567 | 0.4282 | 0.4962 | 0.5214 | 0.4962 | 0.5365 | 0.5416 |
Table 5: The effect of tone on different model for 250 trading days on positive words

| Model          | Schmeling | Kang&Feldman | Tetlock |
|----------------|-----------|--------------|---------|
|                | Fin       | Δ            | Δ + Fin | Δ            | Δ + Fin | Δ            | Δ + Fin |
| Panel A: OEE   |           |              |         |              |         |              |         |
| BNB            | 0.6373    | 0.6324       | 0.6373  | 0.6324       | 0.6422  | 0.6324       | 0.6373  |
| DT             | 0.6373    | 0.5539       | 0.5637  | 0.5539       | 0.5686  | 0.4951       | 0.5735  |
| GNB            | 0.5588    | 0.5588       | 0.5637  | 0.5637       | 0.5686  | 0.5588       | 0.5588  |
| LR             | 0.6373    | 0.6373       | 0.6569  | 0.6373       | 0.6569  | 0.6373       | 0.6569  |
| SVC-linear     | 0.6373    | 0.6373       | 0.6520  | 0.6373       | 0.6520  | 0.6373       | 0.6520  |
| SVC-rbf        | 0.6373    | 0.6373       | 0.6618  | 0.6373       | 0.6618  | 0.6373       | 0.6618  |
| SVC-sigmoid    | 0.6618    | 0.6618       | 0.5294  | 0.6618       | 0.5294  | 0.5882       | 0.5735  |
| Panel B: MPE   |           |              |         |              |         |              |         |
| BNB            | 0.5440    | 0.5440       | 0.6010  | 0.5440       | 0.6114  | 0.5440       | 0.6010  |
| DT             | 0.5440    | 0.5337       | 0.5492  | 0.5337       | 0.5596  | 0.5648       | 0.5440  |
| GNB            | 0.5181    | 0.5026       | 0.5492  | 0.5285       | 0.5544  | 0.5130       | 0.5389  |
| LR             | 0.5440    | 0.5440       | 0.5699  | 0.5492       | 0.5648  | 0.5440       | 0.5855  |
| SVC-linear     | 0.5440    | 0.5440       | 0.5492  | 0.5492       | 0.5492  | 0.5440       | 0.5492  |
| SVC-rbf        | 0.5440    | 0.5492       | 0.5544  | 0.5492       | 0.5544  | 0.5130       | 0.5389  |
| SVC-sigmoid    | 0.5389    | 0.5389       | 0.4611  | 0.5389       | 0.4611  | 0.4767       | 0.5337  |

4. Experimental Results

4.1. The effect of tone change for predicting the trend of stock price

The section analyzes the effect of tone change for predicting the trend of stock price. Although the tone for chairman’s statement section materially differ depending on the computing method and the kind of sentiment words, what perhaps more important is the effect on accuracy for whether the tone is added into input vector and the sensitivity of different tone period. In order to examine the two issues, the work firstly select several commonly models in financial modeling methods used by previous researchers, that is Bernoulli Bayes (BNB), Gaussian Bays (GNB), Decision Tree (DT), Logistic Regression (LR), Support Vector Machine with rbf kernel (SVC-rbf), Support Vector Machine with linear kernel(SVC-linear) and Support Vector Machine with sigmoid kernel (SVC-sigmoid). And then setting up seven labels that use seven tone periods: 3 days, 5 days, 10 days, 20 days, 60 days, 120 days and 250 days. Finally, the paper set up three modeling approaches separately using each machine learning method and each label, one approach that considers each kind of delta tone and financial indicators as input vector respectively (Δt + Fin), second approach that only uses financial indicators as input vector (Fin) and third approach that only selects delta tone as input vector (Δt).

In the experiment, there are three kinds of dataset (ALL, OEE, MPE), each dataset is divided into two section: 75% train set and 25% test set. Table 2, table 3 present the experimental analysis the
effect of delta tone on different trading days for BNB on positive words and negative words. In all tables, Schmeling, Kang&Feldman, Tetlock on table title represents three different computing tone method mentioned before, corresponding to Eq.(5), Eq.(7) and Eq.(10). These tables show that the accuracy (the accuracy is average value of multiple experiments) of 250 trading days tone period is the best on the whole. And the combination of delta tone and financial indicators always get the best result for all kinds of delta tone on all dataset, the highest accuracy is 64.99% for BNB. Besides, among three dataset, OEE dataset get the best result, and result in MPE dataset is worst.

250 trading days is the most sensitive tone period, so the label based on 250 trading days is selected to more in-depth study of the effectiveness of delta tone. Table 4, table 5 represent the effect of delta tone on different models on both positive words and negative words. From these tables, it seems that BNB is the most stable model and the delta tone computed by Kang&Feldman method get the best result.

4.2. Does tone capture other fact information?

The experimental result indicates much more other information. The differences of market efficiency can be seen from table 2. Table 2 shows that the accuracy is the highest on the label on 250 trading days, that is to say, the tone of annual report is more suitable to predict the long-term trend of stock price, but from previous researchers’ works, most research use the short-term abnormal return as a candidate of the prediction label, such as Li et al. [23], Jegadeesh and Wu [15] and so on, in North American Financial Market. So it can make a conclusion that comparing with North American Financial market, Hong Kong Financial market is a relative lower efficiency market.

Figure 1. The accuracy on different year for Financial, Bernoulli Bayes and 250 trading days

The historical financial events’ effect can be seen from experimental result. For example, The Wall Street financial tsunami in 2008 swept the world, resulting in the closure of many fairly large financial institutions or being taken over by the government. Hong Kong financial markets have also been greatly affected. The European debt crisis in 2011 caused global financial market turmoil, the appreciation of the renminbi, the shrinkage of foreign exchange reserves, the limited import and export trade, and the economic downturn in Hong Kong’s financial markets. What’s more, in 2015, China’s introduction of the stock market meltdown mechanism led to further escalation of the stock market disaster and catalyzed the global stock market disaster. Hong Kong and Indian stock markets closed down more than 2 percent. Because the Financial industry is the most sensitive to the change of financial market, the paper select the samples whose company belong to financial industry to further study. Figure 1 represents the accuracy on different year range from 2003 to 2017 for Financial industry. From Figure 2, it is obvious that the effect of these financial events in 2008, 2011, 2015 is that the accuracy decreases obviously because of the sudden financial market changes.

5.Conclusion

This paper explored the sentiment implications of the annual report in Hong Kong market. We collected the annual report of all the listed companies in Hong Kong stock exchange between 2003 and 2017, extracted the companies from Mainland private enterprises(MPE) and overseas-funded and Hong Kong local-funded enterprises(OEE). The knowledge discovery can be summarized as follows.
The paper finds that the tone changes, comparing three different tone changes measures, in the annual report can predict the stock prices in the future. However, the pattern in Hong Kong market is significantly different with the U.S. market, which is more sensitive to the long-term trend for stock price. This evidence proves the market efficiency in Hong Kong market is lower than U.S. The results for Finance sectors show the poor prediction results in the period of the historical financial crisis. Even the bankers, who are widely connected to different economic sectors, cannot foresee the potential financial crisis.

This paper is a pioneer work to investigate the sentiment information of the annual reports in Hong Kong market. There are still lots of rooms to improve. This work only investigates some basic classification models, which build a benchmark for future research. The further research might explore some new classification models, such as multi-kernel SVM and isolated SVM. The current study only focuses on OEE and MPE companies. Future research can put state-owned enterprises, which are going to be the majority in Hong Kong market. However, the characteristics of state-owned enterprises causing the tone changes in the annual reports are different with private enterprise. New sentiment models and linguistic features need to be investigated in the futures.

Acknowledgement

This work was supported by the Shenzhen-Hong Kong Technology Cooperation Fund (SGLH201612091011009 26).

References

[1] Taeyoung Kang, Do Hyung Park, and Ingoo Han. Beyond the numbers: The effect of 10-k tone on firms’ performance predictions using text analytics. Telematics & Informatics, 35(2), 2018.
[2] Kenneth R.French Eugene F.Fama. A five-factor asset pricing model. Journal of Financial Economics, 116:1C22, 2015.
[3] S. R Das. Yahoo! for amazon: Opinion extraction from small talk on the web. In Asia Pacific Finance Association Conference, pages 1375–1388, 2001.
[4] Haifeng You and Xiao Jun Zhang. Financial reporting complexity and investor underreaction to 10-k information. Review of Accounting Studies, 14(4):559–586, 2009.
[5] Wenbin Zhang and Steven Skiena. Trading strategies to exploit blog and news sentiment. In International Conference on Weblogs and Social Media, Icwsm 2010, Washington, Dc, Usa, May, 2010.
[6] Tim Loughran and Bill McDonald. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. Journal of Finance, 66(1):35–65, 2011.
[7] D North T Arnold, RPH Fishe. The effects of ambiguous information on initial and subsequent ipo returns. Financial Management, 39(4):1497C1519, 2010.
[8] S Yan. Managerial attitudes and takeover outcomes: Evidence from corporate filings. Journal of International Financial Markets, Institutions & Money, 35:30C44,2015.
[9] Tamer Elshandidy Yousry Ahmed. The effect of bidder conservatism on m&a decisions:text-based evidence from us 10-k filings. International Review of Financial Analysis, 46:176C190, 2016.
[10] Ronen Feldman, Suresh Govindaraj, Joshua Livnat, and Benjamin Segal. Management’s tone change, post earnings announcement drift and accruals. Review of Accounting Studies, 15(4):915–953, 2010.
[11] Robert P. Schumaker and Hsinchun Chen. Textual analysis of stock market prediction using breaking financial news:the azfin text system. ACM Transactions on Information Systems (TOIS), 27(2):1–19, 2009.
[12] P. Hjek, V. Olej, and R. My?kov. Forecasting stock prices using sentiment information in annual reports - a neural network and support vector regression approach. Wseas Transactions on Business & Economics, 10(4):293–305, 2013.
[13] ChienLiang Chen, ChaoLin Liu, YuanChen Chang, and HsiangPing Tsai. Opinion mining for relating subjective expressions and annual earnings in us financial statements. Journal of Information Science and Engineering, 29(4):743–764, 2012.
[14] ZhiHong Deng, KunHu Luo, and HongLiang Yu. A study of supervised term weighting scheme for sentiment analysis. Expert Systems with Applications, 41(7):3506–3513, 2014.
[15] Narasimhan Jegadeesh and Di Wu. Word power: A new approach for content analysis. Journal of Financial Economics, 110(3):712–729, 2013.
[16] Petr Hjek and Jana Bohov. Predicting abnormal bank stock returns using textual analysis of annual reports c a neural network approach. In International Conference on Engineering Applications of Neural Networks, pages 67–78, 2016.
[17] Victor W. Chu, Raymond K. Wong, Fang Chen, Ivan Ho, and Joe Lee. Enhancing portfolio return based on sentiment-of-topic. Data & Knowledge Engineering, 2017.
[18] Kang C Wang Y, Chen Q. Concurrent sentiment and concept extraction from corporate annual reports for financial performance forecasting. IEEE transactions on smart grid, 7(5):2437–2447, 2015.
[19] R.P.Hart. Redeveloping diction: Theoretical considerations. In Theory, Method, and Practice in Computer Content Analysis, pages 43–60, 2001.
[20] Maik Schmeling and Christian Wagner. Does central bank tone move asset prices? Social Science Electronic Publishing, 2015.
[21] Elaine Henry. Are investors influenced by how earnings press releases are written? Social Science Electronic Publishing, 45(4):363–407, 2006.
[22] McKay Price, James S. Doran, David R Петерсон, and Barbara A. Bliss. Earnings conference calls and stock returns: The incremental informativeness of textual tone. Journal of Banking & Finance, 36(4):992–1011, 2012.
[23] Xiaodong Li, Haoran Xie, Li Chen, Jianping Wang, and Xiaotie Deng. News impact on stock price return via sentiment analysis. Knowledge-Based Systems, 69(1):14–23, 2014.
[24] Ronen Feldman, Suresh Govindaraj, Joshua Livnat, and Benjamin Segal. The incremental information content of tone change in management discussion and analysis. Social Science Electronic Publishing, 108(4):186C191, 2008.
[25] Paul C. Tetlock, Maytal Saar-Tsechansky, and Sofus Macskassy. More than words: Quantifying language to measure firms’ fundamentals. Journal of Finance, 63(3):1437–1467, 2008.