Rule-based Sentiment Analysis for Financial News

Tan Li Im, Phang Wai San, Chin Kim On
Center of Excellence in Semantic Agents
Universiti Malaysia Sabah
Sabah, Malaysia
im_87@hotmail.com, waisanp@hotmail.com,
kimonchin@ums.edu.my

Patricia Anthony
Department of Informatics and Enabling Technologies,
Faculty of Environment, Society and Design
Lincoln University,
Christchurch, New Zealand
patricia.anthony@lincoln.ac.nz

***Abstract*— This paper describes a rule-based sentiment analysis algorithm for polarity classification of financial news articles. The system utilizes a prior polarity lexicon to classify the financial news articles into positive or negative. Sentiment composition rules are used to determine the polarity of each sentence in the news article, while the Positivity/Negativity ratio (P/N ratio) is used to calculate the sentiment values of the overall content of each news article. The performance of the Sentiment Analyser was evaluated using a dataset of manually annotated financial news articles collected from various online financial newspapers. The result was encouraging as our Sentiment Analyser obtained an overall F-Score of 75.6% for both positive and negative classifications.**

Keywords-Sentiment Analysis; Sentiment Composition; Polarity Classification

I. INTRODUCTION

Polarity classification of text can be performed at various levels such as sentence, phrase, word, and document. This paper describes a polarity classification task at document level involving long length text analysis. The aim of this work is to classify financial news articles into positive or negative. Investors often rely on financial news articles to assist them in their investment decision. A sentiment analyser can be used as an investor's tool to quickly classify financial news articles and use this information (along with other information such as the company's profile, how long has it been in business, its financial performance, etc.) to identify potential companies that he can invest in.

The proposed Sentiment Analyser utilizes some existing tools namely the Stanford RNN parser [9] to perform phrase extraction and the Subjectivity lexicon [1] to determine the polarity of the input text. The proposed Sentiment Analyser uses a lexicon-based algorithm which does not involve any machine learning techniques. Instead, a set of sentiment composition rules are created to determine the polarity of the sentences. P/N ratio is incorporated into the proposed Sentiment Analyser to determine the polarity of the overall content of each financial news articles.

In this paper, we present the general framework of the proposed Sentiment Analyser, and a brief description of its relevant components. Additionally, the sentiment composition rules used in the Sentiment Analyser are described. We carried out experimental evaluation to test the performance of the proposed Sentiment Analyser in classifying financial news articles and the results are elaborated in this paper.

II. LITERATURE REVIEWS

Sentiment analysis is a field in natural language processing with the aims of trying to figure out what other people think towards a topic of interest by using computational power [2, 3]. Sentiment analysis is useful in pinpointing the sentiment of the resources automatically without the need of excess manpower to go through the long and winded document which sometimes may only contain a few sentences that convey the thoughts of the author [4].

Newspaper is a credible platform to publish new information to the public. It is among the most reliable way of conveying the information via text compared to other platforms such as rumors, scandal, and eavesdropping [3]. Today, many researches in sentiment analysis are rooted in examining the relationship of financial news articles to the stock market behaviour. For example, the works of Daniel [5] and Schumaker et al. [6] focused on sentiment analysis using financial news articles.

There are various approaches that can be used to perform sentiment analysis. According to Annett and Kondrak [12], Taboada and Brook [11], sentiment analysis can be categorized into two main approaches which are lexicon-based approaches and machine learning approaches. Lexicon-based approaches utilize the prior polarity lexicon to determine the semantic orientation of the document, while machine learning approaches typically use the classifier to classify documents according to its semantic orientation.

Sentiment composition is one of the techniques used to solve the problem in sentiment analysis. The sentiment composition techniques interpret the analysed text by its compositional structure. These techniques can be used to perform polarity classification up to sentence level. Klenner's work [7, 8] is one of the examples that used sentiment composition to perform polarity classification. His work is focused on the extraction of noun phrase detection using pattern-matching sentiment composition rules. The results obtained from the experiments were encouraging as they showed that sentiment composition can be used to solve the polarity classification problem. Another work from [10], utilized compositional approach to sentiment analysis. They introduced the quasi-compositionality to perform sentiment

978-1-4799-8697-2/15 $31.00 © 2015 IEEE
DOI 10.1109/SMC.2015.283

classification and showed that it was able to solve the sentiment propagation's problem, polarity reversal, and sentiment sense disambiguation.

III. The Sentiment Analyser

This section describes the general framework of the proposed Sentiment Analyser. The proposed work is based on Klenner's work [7]. In his work, Klenner used sentiment composition to perform sentiment classification at sentence level. We modified and extended his algorithm to accommodate sentiment classification at document level. However, instead of using the original words from the analysed text, we changed the representation of each word to lemma representation. In addition, we added new sentiment composition rules.

The proposed Sentiment Analyser utilized the Stanford RNN parser [9] to perform phrase extraction. A prior polarity lexicon namely Subjectivity Lexicon [1] is used as a guideline for the polarity tagging task in the system. The sentiment composition rules are applied to determine the polarity of each sentence in the financial news article. The P/N ratio is used to calculate the sentiment value of the overall content of the financial news article.

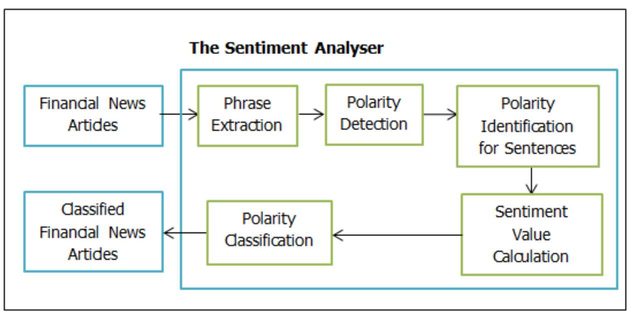

Figure 1. *Framework for the Sentiment Analyser*

Fig. 1 shows the framework for the proposed Sentiment Analyser. As shown in Fig. 1, the Sentiment Analyser consists of five main phases that include the phrase extraction phase, the polarity detection phase, the polarity identification phase, the sentiment value calculation phase, and the polarity classification phase. The end result is the classification of the financial news article into positive or negative.

A. Phrase Extraction

The phrase extraction phase takes the financial news article as input and splits the input into sentences. The extracted sentences are then parsed according to its Part of Speech (POS) tags using Stanford RNN Parser [9].

B. Polarity Detection

The polarity detection phase deals with lexicon matching and polarity tagging of the parsed text obtained from the earlier phase. In this work, the subjectivity lexicon (Wilson, 2005), which is a prior polarity lexicon is used as a backbone to support the polarity identification of the words in the analysed text. Each of the word in the analysed text is matched with the clues in the subjectivity lexicon. If a matched pair is found in the lexicon, then the polarity of the word will be tagged according to its matched pair's polarity.

C. Polarity Identification for Sentences

In the polarity identification phase, the sentiment composition rules are applied. The sentiment composition is written in a rule-based format. The rules are tailor made to suit the output generated from the phrase extraction phase which is POS tagged using the Penn Treebank - Penn POS tags. The sequences of rule are important in this task to ensure the correctness of the interpretation of the meaning of the phrase/sentence. These rules go from a simple form to a complicated form in a sequential process. The identification of the polarity of the sentences is considered to be completed when all the rules are applied to the extracted phrase which has been polarity tagged.

D. Sentiment Value Calculation

This phase calculates the sentiment value of the entire article. The sentiment value is calculated using the mathematical formula called P/N ratio where it uses the number of positive sentences and negative sentences obtained from the sentence polarity identification task. The sentiment value of the financial news article is calculated by averaging the sentiment values of all the sentences in the financial news article.

E. Polarity Classification

The financial news article is classified based on the overall sentiment value calculated using the P/N ratio. In this work, we focus on classifying the financial news articles into two categories, which are positive and negative. We used +1 to indicate the greatest positivity sentiment value and -1 to indicate greatest negativity sentiment value of the financial news article. The value of 0 is obtained when the news article is neutral or there is no sentiment detected in the analysed news (this value is ignored in our case).

IV. Usage of the Sentiment Composition Rules

The sentiment composition rules is a set of rules which is applied to determine the polarity of the sentences. This set of sentiment composition rules is created to tackle noun phrase sentiment composition, preposition phrase sentiment composition, and phrase to phrase sentiment composition based on Klenner's work [7]. In addition, we added new sentiment composition rules to tackle verb phrase sentiment composition, verb-noun/noun-verb phrase sentiment composition, the conjunction "*but*" sentiment composition, and the negation.

A. Noun Phrase Sentiment Composition Rules

Noun phrase sentiment composition rules covered combinations of an adjective merged with noun (JJ-NN), JJ adjective merged with noun phrase (JJ-NP), an adjective phrase merged with noun (ADJP-NN), and an adjective phrase merged with noun phrase (ADJP-NP). These four combinations form a noun phrase composition. Table I shows the sentiment composition rules for noun phrase combinations.

TABLE I. NOUN PHRASE SENTIMENT COMPOSITION RULES

Rules	*POS Combination(JJ-NN/ JJ-NP/ ADJP-NN/ ADJP-NP)*	→	*Output (NP)*
NP1	(NEG)(POS)	→	NEG
NP2	(NEG)(NEG)	→	NEG
NP3	(NEG)(NEU)	→	NEG
NP4	(POS)(POS)	→	POS
NP5	(POS)(NEG)	→	NEG
NP6	(POS)(NEU)	→	POS
NP7	(NEU)(POS)	→	POS
NP8	(NEU)(NEG)	→	NEG

Consider the rule NP1 in Table I. A combination of negative (NEG) JJ/ADJP and a positive (POS) NN/NP will yield an output of a negative noun phrase (NP). For instance, the negative adjective phrase "*most difficult*" when merged with the positive noun phrase "*business decision*" will result in a negative noun phrase of "*most difficult business decision*".

B. Verb Phrase Sentiment Composition Rules

Verb phrase sentiment composition rules involve the combinations of an adjective merged with verb (JJ-VB), an adjective merged with verb phrase (JJ-VP), an adjective phrase merged with verb (ADJP-VB), and an adjective phrase merged with verb phrase (ADJP-VP). These combinations form a verb phrase composition. Table II shows the verb phrase sentiment composition rules for the combinations of JJ-VB, JJ-VP, ADJP-VB, and ADJP-VP.

TABLE II. VERB PHRASE SENTIMENT COMPOSITION RULES

Rules	*POS Combination(JJ-VB/ JJ-VP/ ADJP-VB/ ADJP-VP)*	→	*Output (VP)*
VP1	(NEG)(POS)	→	NEG
VP2	(NEG)(NEG)	→	NEG
VP3	(NEG)(NEU)	→	NEG
VP4	(POS)(POS)	→	POS
VP5	(POS)(NEG)	→	NEG
VP6	(POS)(NEU)	→	POS
VP7	(NEU)(POS)	→	POS
VP8	(NEU)(NEG)	→	NEG

Consider rule VP4 in Table II; a positive JJ/ADJP combined with a positive VB/VP will form a positive verb phrase. For example, the positive adjective "*top*" combined with positive verb "*preferred*" will yield a positive verb phrase "*top preferred*".

C. Verb-Noun/Noun-Verb Phrase Sentiment Composition Rules

The sentence pattern in this categories includes the combination of a verb merged with noun (VB-NN), verb merged with noun phrase (VB-NP), verb phrase merged with noun (VP-NN), verb phrase merged with noun phrase (VP-NP), noun merged with verb (NN-VB), noun merged with verb phrase (NN-VP), noun phrase merged with verb (NP-VB), and noun phrase merged with verb phrase (NP-VP).

TABLE III. VERB-NOUN/ NOUN VERB PHRASE SENTIMENT COMPOSITION RULES

Rules	*POS Combination(VB-NN/ VB-NP/VP-NN/VP-NP/ NN-VB/NN-VP/ NP-VB/NP-VP)*	→	*Output (phrase/sentence)*
VN1	(NEG)(POS)	→	NEG
VN2	(NEG)(NEG)	→	NEG
VN3	(NEG)(NEU)	→	NEG
VN4	(POS)(POS)	→	POS
VN5	(POS)(NEG)	→	NEG
VN6	(POS)(NEU)	→	POS
VN7	(NEU)(POS)	→	POS
VN8	(NEU)(NEG)	→	NEG

Table III shows the sentiment composition rules for verb-noun/noun-verb composition. As shown in rule VN5, when a positive noun phrase merges with a negative verb phrase it will produce a negative phrase or sentence. For example, the phrase "*gains from asset sales dwindled*" is the combination of a positive noun phrase with a negative verb phrase which results in a negative noun phrase.

D. Preposition Phrase Sentiment Composition Rules

The preposition is a word that combines two or more phrases together to form a new phrase or sentence. The most commonly used preposition in English includes "*in*", "*on*", "*to*", "*of*", "*by*" and so on. In this work, the sentiment composition rules are set to cover the preposition of "*in*", "*to*", and "*of*". Table IV shows the sentiment composition rules for phrase to phrase combination with the preposition "*in*".

TABLE IV. PHRASE-PHRASE SENTIMENT COMPOSITION RULES WITH PREPOSITION "IN"

Rules	*POS Combination(phrase "in" phrase)*	→	*Output (phrase/sentence)*
Pin1	(NEG (PP(IN)POS))	→	NEG
Pin2	(NEG (PP(IN)NEG))	→	NEG
Pin3	(NEG (PP(IN)NEU))	→	NEG
Pin4	(POS (PP(IN)POS))	→	POS
Pin5	(POS (PP(IN)NEG))	→	POS
Pin6	(POS (PP(IN)NEU))	→	POS
Pin7	(NEU (PP(IN)POS))	→	POS
Pin8	(NEU (PP(IN)NEG))	→	NEG

The rule Pin5 shows that when a positive phrase merges with another negative phrase with the preposition "*in*", the output will be a positive phrase or sentence. The phrase "*solid performance in a seasonally slow quarter*" is an example that matches Pin5. These rules are applicable in situation where a phrase is combined with another phrase with the preposition "*on*", "*by*", "*as*", "*that*", and "*for*" because they share the same patterns as the preposition "*in*".

Table V shows the sentiment composition rules for phrase to phrase composition with the preposition "*to*".

TABLE V. PHRASE-PHRASE SENTIMENT COMPOSITION RULES WITH PREPOSITION "TO"

Rules	*POS Combination(phrase "to" phrase)*	→	*Output (phrase/sentence)*
Pto1	(NEG (PP(TO)POS))	→	NEG
Pto2	(NEG (PP(TO)NEG))	→	NEG
Pto3	(NEG (PP(TO)NEU))	→	NEG
Pto4	(POS (PP(TO)POS))	→	POS
Pto5	(POS (PP(TO)NEG))	→	POS
Pto6	(POS (PP(TO)NEU))	→	POS
Pto7	(NEU (PP(TO)POS))	→	POS
Pto8	(NEU (PP(TO)NEG))	→	NEG

Based on rule Pto1, a negative phrase combined with a positive phrase with the preposition "*to*" produces a negative phrase or sentence. For instance, the phrase "*serious problem to the management*" is an example of Pto1.

Table VI shows the sentiment composition rules for the combination of phrase to phrase with the preposition "*of*".

TABLE VI. PHRASE-PHRASE SENTIMENT COMPOSITION RULES WITH PREPOSITION "OF"

Rules	*POS Combination(phrase "of" phrase)*	→	*Output (phrase/sentence)*
Pof1	(NEG (PP(OF)POS))	→	NEG
Pof2	(NEG (PP(OF)NEG))	→	NEG
Pof3	(NEG (PP(OF)NEU))	→	NEG
Pof4	(POS (PP(OF)POS))	→	POS
Pof5	(POS (PP(OF)NEG))	→	POS
Pof6	(POS (PP(OF)NEU))	→	POS
Pof7	(NEU (PP(OF)POS))	→	POS
Pof8	(NEU (PP(OF)NEG))	→	NEG

In rule Pof2, a negative phrase is combined with a negative phrase by the preposition "*of*" to form a negative phrase or sentence. An example of the Pof2 rule is the phrase "*fear of losing control*".

E. The Conjunction "but" Sentiment Composition Rules

The conjunction "*but*" is used to combine two phrases/sentences into one. Table VII shows the sentiment composition rules for the conjunction "*but*".

TABLE VII. THE CONJUNCTION "BUT" SENTIMENT COMPOSITION RULES

Rules	*POS Combination(phrase "but" phrase)*	→	*Output (phrase/sentence)*
Cbut1	NEG (BUT) POS	→	POS
Cbut2	NEG (BUT) NEG	→	NEG
Cbut3	NEG (BUT) NEU	→	NEG
Cbut4	POS (BUT) POS	→	POS
Cbut5	POS (BUT) NEG	→	NEG
Cbut6	POS (BUT) NEU	→	NEG
Cbut7	NEU (BUT) POS	→	POS
Cbut8	NEU (BUT) NEG	→	NEG

The rule Cbut1 in Table VII shows that when a negative phrase/sentence is combined with positive phrase/sentence, the output will be a positive phrase/sentence. For example, consider the sentence "*The first half of the year was a nightmare but we have performed much better in the second half of the year*." The first half of the sentence is negative while the second half is positive, and the output of this sentence is a positive sentence.

F. Negation Rules

Negation is a polarity shifter that turns a positive statement into negative or vice versa. It plays a crucial role in the linguistic structure of a sentence. Adding different polarity shifters to a sentence will result in a different opinion towards the same topic. For instance, consider the sentence below:

I like the movie.
I don't like the movie.
I deeply like the movie.
I rather like the movie.

All the sentences shown above are discussing the same topic - the "*movie*". With the presence of the different polarity shifters, each of the sentences shows different degrees of "*like*" towards the movie. The polarity shifter "*don't*" in the second sentence flips the polarity of "*like*" from positive to negative "*not like*". The third sentence "*deeply*" intensifies the "*like*" to "*love*", while the term "*rather*" decreases the "*like*" to "*little like*".

The term that acts as the polarity shifter such as "*no*", "*not*", "*never*", "*neither*", "*none*", "*without*", "*below*", and so on are also included in this work. Table VIII shows the negation sentiment composition rules. There are three rules in this category which involve the negation of positive phrase, the negation of negative phrase, and the negation of neutral phrase.

TABLE VIII. THE NEGATION SENTIMENT COMPOSITION RULES

Rules	***POS Combination(NGT(phrase))***	→	***Output (phrase/sentence)***
NGT1	(NGT)NEG	→	POS
NGT2	(NGT)POS	→	NEG
NGT3	(NGT)NEU	→	NEU

As shown in rule NGT1 in Table VIII, a negation of a negative phrase will result in a positive output. For example, the term "*scary*" is negative. If a polarity shifter "*not*" exists in front of this term, then the output "*not scary*" will be positive.

V. USING THE SENTIMENT ANALYSER

In this section, we present the step by step process of the sentiment analyser in analyzing the financial news articles. Consider the example of a snippet taken from one of the financial news article given below:

"*Hong Leong Bank Bhd's pre-tax profit surged 45 per cent to RM638 million. Total income rose 0.5 per cent to RM1 billion, Group Managing Director Datuk Yvonne Chia said. We continued to achieve another set of strong results in this seasonally slow quarter.*"

From the example, the input will be split into three sentences and parsed using the Stanford RNN parser. Fig. 2 below shows the parsed text for the respective sentence.

```
Hong Leong Bank Bhd's pre-tax profit surged 45 per cent to RM638 million.
(ROOT(S(NP (NP (NNP Hong) (NNP Leong) (NNP Bank) (NNP Bhd) (POS 's))
 (JJ pre-tax) (NN profit))(VP (VBD surged)(NP(NP (CD 45))(PP (IN per)
(NP (NN cent))))(PP (TO to) (NP(QP (CD RM638) (CD million)))))(. .)))

Total income rose 0.5 per cent to RM1 billion, Group Managing Director Datuk Yvonne Chia said.
(ROOT(S(S (NP (JJ Total) (NN income))(VP (VBD rose)(NP(NP (CD 0.5))
(PP (IN per)(NP (NN cent))))(PP (TO to)(NP(QP (CD RM1) (CD billion))))))(, ,)
(NP (NNP Group) (NNP Managing) (NNP Director) (NNP Datuk) (NNP Yvonne)
 (NNP Chia))(VP (VBD said))(. .)))

We continued to achieve another set of strong results in this seasonally quarter."
(ROOT (S (NP (PRP We)) (VP (VBD continued)(S (VP (TO to)
(VP (VB achieve)(NP(NP (DT another) (NN set))(PP (IN of)(NP (JJ strong)
(NNS results))))(PP (IN in) (NP (DT this) (ADJP (RB seasonally)(JJ slow))(NN quarter))))))(. .) ('' '')))
```

Figure 2. *Parsed Text*

Once the sentences are parsed according to its POS, we perform polarity tagging by matching the words in the analysed text with the prior polarity lexicon. Fig. 3 shows the example of polarity detection.

```
Hong Leong Bank Bhd's pre-tax profit surged 45 per cent to RM638 million.
(ROOT(S(NP (NP (NNP Hong) (NNP Leong) (NNP Bank) (NNP Bhd) (POS 's))
 (JJ pre-tax) (NN profit))(VP (VBD surged)(NP(NP (CD 45))(PP (IN per)
(NP (NN cent))))(PP (TO to) (NP(QP (CD RM638) (CD million)))))(. .)))

Total income rose 0.5 per cent to RM1 billion, Group Managing Director Datuk Yvonne Chia said.
(ROOT(S(S (NP (JJ Total) (NN income))(VP (VBD rose)(NP(NP (CD 0.5))
(PP (IN per)(NP (NN cent))))(PP (TO to)(NP(QP (CD RM1) (CD billion))))))(, ,)
(NP (NNP Group) (NNP Managing) (NNP Director) (NNP Datuk) (NNP Yvonne)
 (NNP Chia))(VP (VBD said))(. .)))

We continued to achieve another set of strong results in this seasonally slow quarter."
(ROOT (S (NP (PRP We)) (VP (VBD continued)(S (VP (TO to)
(VP (VB achieve)(NP(NP (DT another) (NN set))(PP (IN of)(NP (JJ strong)
(NNS results))))(PP (IN in) (NP (DT this) (ADJP (RB seasonally)(JJ slow))(NN quarter))))))(. .) ('' '')))
```

Figure 3. *Polarity Detection*

Based on Fig. 3, we can see that the bold words in green indicates that they are words that carry a positive sentiment, the bold words in red are words with negative sentiment and the bold words in blue are words with neutral sentiment. Next, we applied the sentiment composition rules to each of the sentences. Fig. 4 shows how the composition rules are applied to identify the polarity of the sentence.

```
Hong Leong Bank Bhd's pre-tax profit surged 45 per cent to RM638 million.
(ROOT(S(NP (NP (NNP) (NNP) (NNP) (NNP) (POS 's))
 (JJ) (NN POS))(VP (VBD POS)(NP(NP (CD))(PP (IN)
(NP (NN))))(PP (TO) (NP (QP (CD ) (CD)))))(. .)))

(ROOT(S(NP POS)(VP POS)))                                        Apply rule VN4

(ROOT POS)

Total income rose 0.5 per cent to RM1 billion, Group Managing Director Datuk Yvonne Chia said.
(ROOT(S(S (NP (JJ) (NN))(VP (VBD POS)(NP(NP (CD))
(PP (IN)(NP (NN))))(PP (TO)(NP(QP (CD) (CD))))))(, ,)
(NP (NNP) (NNP) (NNP) (NNP) (NNP)
 (NNP))(VP (VBD))(. .)))

(ROOT(S(VP POS))

(ROOT POS)

We continued to achieve another set of strong results in this seasonally slow quarter.
(ROOT (S (NP (PRP)) (VP (VBD POS)(S (VP (TO)
(VP (VB POS)(NP(NP (DT) (NN))(PP (IN)(NP (JJ POS)
(NNS NEU))))(PP (IN) (NP (DT) (ADJP (RB)(JJ NEG))(NN))))))(. .) ('' '')))

(ROOT (S(VP POS)(S (VP (TO to)(VP POS)(NP (JJ POS) (NNS NEU)))) ))))(PP (IN) (NP NEG) Apply rule NP6

(ROOT (S(VP POS)(S (VP (TO to)(VP POS)(NP POS))))(PP (IN) (NP NEG)    Apply rule VN4

(ROOT (S(VP POS)(S (VP (TO to)(VP POS)) (PP (IN) (NP NEG)             Apply rule Pin5

(ROOT (S(VP POS)(S (VP (TO to)(VP POS)))                             Apply rule Pto4

(ROOT (S(VP POS)))

(ROOT POS)
```

Figure 4. *Polarity Identification for sentences*

The first sentence is matched with sentiment rule VN4 which states that a (NP POS)(VP POS) yields POS. Therefore, this sentence is positive. The second sentence matches the sentiment composition rule VN4 with (VP POS)(NP POS) →POS, which indicates that this sentence is positive as well. The third sentence matches the sentiment composition rule NP6 where (JJ POS)(NN NEU) →POS. Next, it performs rule VN4 which indicates that (VP POS)(NP POS) →POS. After that, it applies rule Pin5 where (POS(PP(IN)NEG)) →POS. Finally, the last rule, Pto4 ((POS (PP(TO)POS)) →POS) evaluates to positive**.** Therefore, the third sentence is identified by the system as positive. After the polarity identification process, we calculate the overall sentiment value of the analysed text. As all three sentences in this example are positive, it is assigned a value of 1 to indicate that it has a greatest positivity. Based on the sentiment value obtained, the sentiment analyser classifies this financial news article as positive.

VI. EXPERIMENT AND DISCUSSION

A. Dataset

The dataset used in the experiment is collected from various Malaysian online newspapers including The National News Agency of Malaysia (BERNAMA), The Sun Daily, The Star Online, and The New Strait Times (NST). The collected news articles were manually annotated as positive or negative by two different annotators and validated through the interannotator agreement. A total of 200 financial news articles were used in the experiment described in this paper.

B. Performance Metrics

The performance of the proposed Sentiment Analyser is measured using the Recall, Precision and F-Score. In this work, we measured the performance of the Sentiment Analyser into three cases, positive case classification, negative case classification, and all cases classification. Positive case classification measures the ability of the proposed Sentiment Analyser in classifying positive news articles, while negative case classification measures the ability of classifying negative news articles, and finally the all cases classification measures the system's ability to classify the financial news articles into positive or negative in overall. We compare the performance of our proposed Sentiment Analyser against the baseline sentiment analyser. The baseline sentiment analyser is very similar to Klenner's work but we modified it to perform classification at document level and tested it with our own dataset. It uses sentiment composition rules that cover noun phrase sentiment composition and phrase to phrase sentiment composition with preposition "*in*", "*to*", and "*of*".

C. Experiment Results

We ran both the baseline and the proposed Sentiment Analyser with the financial news articles dataset as input. Table IX shows the results obtained by the baseline sentiment analyser and the proposed Sentiment Analyser in positive case classification, negative case classification, and all cases classification.

TABLE IX. THE COMPARISON OF PERFORMANCE OF THE BASELINE SENTIMENT ANALYSER TO THE PROPOSED SENTIMENT ANALYSER

Measurement in %	***Baseline***			***Sentiment Analyser***		
	Positive Case	***Negative Case***	***All Cases***	***Positive Case***	***Negative Case***	***All Cases***
Recall	85.8	25.0	61.5	88.3	50.0	**73.0**
Precision	67.3	66.7	67.2	76.3	85.1	**78.5**
F-Score	75.5	36.4	64.2	81.9	63.0	**75.6**

The Sentiment Analyser performed well in positive case classification with recall score of 88.3%, precision of 76.3% and F-Score of 81.9%. This is an average improvement of more than 8% when compared to the result obtained by the baseline sentiment analyser. The proposed Sentiment Analyser also outperformed the baseline sentiment analyser in both the negative classification and all cases classification. The lower F-Score for the negative case classification indicates that our Sentiment Analyser's algorithm is positive bias. However, the results obtained showed that the proposed Sentiment Analyser has improved in performance even in negative case classification with an F-score of 63.0% compared to the F-Score of 36.4% obtained by the baseline sentiment analyser. In all cases classification, the Sentiment Analyser recorded an F-Score of 75.6% which shows an improvement of 11.4% from the baseline system which obtained an F-score of 64.2%.

VII. CONCLUSION AND FUTURE WORK.

This paper describes the rules-based sentiment analysis algorithms with the aid of prior polarity lexicon in performing polarity classification for financial news articles. We proposed the used of sentiment composition rules together with the P/N ratio to identify the polarity of the financial news articles. Based on the results that we obtained from the experiment, the proposed system scored an average F-Score of 75.6% in all classifications. This is much better compared to the baseline sentiment analyser which obtained an F-Score of 64.2%. There are some limitations in the proposed Sentiment Analyser. Firstly, the sequences of the rules are fixed and cannot be changed. In its current form, the sentiment composition rules must go from simple composition rules to more complicated composition rules in a sequential process to avoid misinterpretation of the meaning of the phrase or sentence. Secondly, our proposed sentiment analyser is not able to tackle the word ambiguity problem where there are words in the subjectivity lexicon which are tagged with multiple polarities. This can lead to misclassifications of sentence's polarity which also affects the misclassification of articles. For instance, the phrase "*cost reduction*" which matches the rule VN2 is misclassified as negative. In the future, we will consider using semantic similarity techniques to perform sentiment enrichment, as well as to solve the words' ambiguity problem.

REFERENCES

[1] Wilson, Theresa, J. Wiebe, and P. Hoffmann. "Recognizing contextual polarity in phrase-level sentiment analysis." In *Proceedings of the conference on human language technology and empirical methods in natural language processing*, pp. 347-354. Association for Computational Linguistics, 2005.

[2] P. Bo, and L. Lee. "Opinion mining and sentiment analysis." *Foundations and trends in information retrieval* 2, no. 1-2, 2008: 1-135.

[3] K. Wiltrud. "Introduction to Sentiment Analysis,", 2012.

[4] H. Minqing, and B. Liu. "Mining and summarizing customer reviews." In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 168-177. ACM, 2004.

[5] A. Pablo Daniel. "Sentiment analysis in financial news." PhD diss., Harvard University, 2009.

[6] Schumaker, Robert P., Y. Zhang, C.N. Huang, and H. Chen. "Evaluating sentiment in financial news articles." Decision Support Systems 53, no. 3 (2012): 458-464.

[7] K. Manfred, S. Petrakis, and A. Fahrni. "Robust Compositional Polarity Classification." In RANLP, pp. 180-184. 2009.

[8] K. Manfred, A. Fahrni, and S. Petrakis. "PolArt: A robust tool for sentiment analysis." In Proceedings of the 17th Nordic Conference of Computational Linguistics, vol. 4, pp. 235-238. 2009.

[9] K. Lingpeng, and Noah A. Smith. "An empirical comparison of parsing methods for stanford dependencies." arXiv preprint arXiv:1404.4314, 2014.

[10] M. Karo, and S. Pulman. "Sentiment composition." In Proceedings of the Recent Advances in Natural Language Processing International Conference, pp. 378-382. 2007.

[11] Taboada, Maite, J. Brooke, M. Tofiloski, K. Voll, and M. Stede. "Lexicon-based methods for sentiment analysis." Computational linguistics 37, no. 2 2011, pp.267-307.

[12] Annett, Michelle, and G. Kondrak. "A comparison of sentiment analysis techniques: Polarizing movie blogs." In Advances in artificial intelligence, pp. 25-35. Springer Berlin Heidelberg, 2008.