Thai Stock News Sentiment Classification using Wordpair Features

Apinan Chattupan
Knowledge Management and Knowledge Engineering Laboratory (KMAKE Lab)
Faculty of Information Technology
King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand
S7606151@kmitl.ac.th

Ponrudee Netisopakul
Knowledge Management and Knowledge Engineering Laboratory (KMAKE Lab)
Faculty of Information Technology
King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand
ponrudee@it.kmitl.ac.th

Abstract

Thai stock brokers issue daily stock news for their customers. One broker labels these news with plus, minus and zero sign to indicate the type of recommendation. This paper proposed to classify Thai stock news by extracting important texts from the news. The extracted text is in a form of a 'wordpair'. Three wordpair sets, manual wordpairs extraction (ME), manual wordpairs addition (MA), and automate wordpairs combination (AC), are constructed and compared for their precision, recall and f-measure. Using this broker's news as a training set and unseen stock news from other brokers as a testing set, the experiment shows that all three sets have similar results for the training set but the second and the third set have better classification results in classifying stock news from unseen brokers.

Keywords: Thai stock news, sentiment classification, text classification, wordpair features.

1 Introduction

Thai stock news are daily issued from many stock brokers. Thai stock news is an important source of information for stock traders to make a decision on stock trading. However, a usual Thai stock news has a long message and sometimes not easily to interpret or conclude. One stock broker makes it easier by labeling each news with plus (+), minus (-) and zero (0) sign, to indicate the type of news as positive, negative and neutral. This automatically classifies the news into three classes.

In this research, we assume that 'features' that can be used for classifying the news must be presented as text in the news. Although this assumption could be too strong in general, for our sole purpose of investigation, our focus here is on text form of the news. Therefore, we proposed to construct a set of these 'texts' to be used as features in order to classify Thai stock news into three sentiments: positive, negative and neutral classes, using known sentiment news as a training set and unseen news as a testing set.

Each feature is called a *wordpair*, since it is a pair of a keyword and a polarity word. A keyword is a word that signifies upcoming information. A polarity word is a word associated with a keyword and signifies a sentiment that related to the keyword. Following the classification from one broker, there are three sentiments: positive, negative, and neutral.

However, due to the flexibility of Thai language, the order of a keyword, a polarity word and a stock symbol may not be the same in the news. That is - a keyword may come before or after a polarity word. In addition, they may come before, after or between the stock symbols they intend to recommend.

29th Pacific Asia Conference on Language, Information and Computation pages 188 - 195
Shanghai, China, October 30 - November 1, 2015

There are two objectives of this paper. First, describing methods to construct these wordpairs collection. Second, utilizing them for constructing automatic classification models for classifying Thai stock news into three corresponding classes. This method can be very useful for general investors because investors can quickly obtain the information and make a decision in stock trading by following the trend of Thai stock news from the classification model.

The outline of this paper is as follows. Section 2 reviews related work. Section 3 describes stock news collection and wordpair construction. In subsection 3.1, we show an example stock news, signs, and stock symbols and compare the frequency of stock symbols in the training and testing set. In subsection 3.2, we propose three sets of wordpairs, which are used to classify Thai stock news sentiments. Section 4 gives details of two experimental designs. We also discuss an effect of varying window sizes for extracting wordpairs features. The results of stock news sentiment classification are also shown in this section. Section 5 analyzes misclassified stock news from the testing set. The last section gives a conclusion and our plan for imminent future work.

2 Related Work

There are two involving areas related to our work. First, the research involved language structure and processing research. Second, the research involved analyzing the stock and classification.

Tongtep and Theeramunkong (2010) mentioned the structural model for extracting patterns from Thai news documents. They focus on a pattern of unique name or noun such as person name, organization name, location, date and time. In addition, Sutheebanjard and Premchaiswadi (2010); Lertcheve and Aroonmanakun (2009) mentioned a similar extracting pattern. They extracted only the person name and only the product name respectively.

Taboada et al. (2011); Lertsuksakda et al. (2014) mentioned the types of word; such as a noun, a verb, an adjective and adverb; and polarity of the word. Taboada et al. (2011) discussed the types of word that have an emotional level and the negation of word that will affect to an emotional level. Lertsuksakda et al. (2014) discussed Thai sentiment terms by using the hourglass of emotion. They assigned an emotional level of a Thai word using two-way translation from English word corpus to Thai word. These techniques will be applied to extract wordpairs from Thai stock news.

Mittermayer (2004); Schumaker and Chen (2009); Chattupan and Netisopakul (2014) demonstrated stock trends prediction using text in the stock news. In addition, Lertsuksakda et al. (2015) discussed text mining techniques in Thai children stories. We will take above techniques and apply them to our work.

3 Stock News and Wordpairs

Section 3.1 describes data preparation including stock news collection and preliminary analysis. Section 3.2 describes wordpairs construction for stock news sentiment classification experiments.

3.1 Stock News Collection

The experiment collects Thai stock news from several brokers such as Bualuang securities (BLS), Thanachart securities (TNS), Krungsri securities (KSS), and so on. In this paper, we tag wordpairs from BLS stock news, hence, we use news from this set as a training set. Stock news from other brokers are combined and used as a testing set. Another important reason for using BLS as a training set is that the broker recommendation includes sentiment signs, such as +, 0, -, *. Therefore, wordpairs sentiments from this broker can be directly tagged using the sign sentiment. The example of stock news published by BLS with their signs are shown in Table 1. Note that some news contain more than one stock symbols.

Thai stock news from BLS (Bualuang Securities, 2015) was collected between 04/04/2014 to 27/05/2015. There are 1,381 stock news with totally 6,596 paragraph news containing stock symbols. Paragraph length is from 200 to 500 letters. Furthermore, Thai stock news under investigation was selected with the following condition. First, the selected stock news must have a stock symbol. This is used for obtaining stock statistics; percent price changes and trading volume. Second, stock news must contain at least one wordpairs. Hence, the stock news contains only a stock symbol and recommended price will not be selected.

Date	Stock news	Sign	Symbol
17/02/2015	MAKRO กำไรดีกว่าคาด คงกำไรปีนี้ / คงคำแนะนำถือ makro-kum-rai-dee-kwa-kard-kong-kum-rai-pee-nee-/-kong-kum-nae-num-thux 'Makro earn better than expected. This year continued profit. / Recommend hold.'	0	MAKRO
06/10/2014	กลุ่มโรงแรมท่องเที่ยวรายงาน... เรายังคงแนะนำซื้อ MINT CENTEL และ ERW koom-rong-ram-tong-tiew-rai-ngan-…-rao-young-kong-nae-num-shux-mint-centel-lae-erw 'The hotel and travel report… We recommend buy Mint, Centel and Erw.'	+	MINT CENTEL ERW
17/02/2015	เช้านี้เกาหลีใต้ คงดอกเบี้ย 2% Shao-nee-kao-lee-tai-kong-dok-bia-song-per-cent 'This morning, south korea fixed interest rate 2%.'	+	No

Table 1: The example of Thai stock news published by BLS

Thai stock news from other brokers (Stock News Online, 2015) were collected between 10/03/2015 to 03/07/2015 and has totally 3,489 paragraphs news. We used the same selecting criteria as the training set. However, we notice that this set has longer paragraph length from 500 to 1000 letters. We prepare this testing set from unseen data set in order to support for future unseen stock news.

Industry	BLS		Other Brokers	
	Freq.	Rank	Freq.	Rank
Agro	780	3	303	5
Consump	310	8	139	8
Fincial	514	5	345	3
Indus	359	7	189	7
Propcon	2512	1	1356	1
Resourc	629	4	265	6
Service	1094	2	562	2
Tech	398	6	330	4

Table 2: Comparing frequency of stock symbols grouped by types of industry in training and testing set

Table 2 shows the preliminary investigation of stock symbols frequencies grouped by types of industry, comparing the training set (BLS) and the testing set (other brokers). Notice that for both sets, the most and second most frequent stock symbols are in industry: Propcon and Service; while the least and second least frequent symbols are in industry: Consump and Indus, respectively. The total number of unique symbols in the training set is 296 comparing to 219 in the testing set. Hence, the average mentioned frequency for each symbol in the training and testing set are 51.74% and 38.28%, respectively. The total numbers of symbols grouped by types of industry are shown in Table 3.

Industry	Unique symbols in SET	Unique symbols in training set	Unique symbols in testing set
Agro	54	32	29
Consump	42	12	8
Fincial	61	32	27
Indus	87	32	17
Propcon	152	76	53
Resourc	36	28	20
Service	99	57	41
Tech	41	27	23
Summary	572	296	219

Table 3: The total number of unique symbols grouped by types of industry in SET, training set, and testing set

3.2 Wordpairs Construction

This paper proposed to use stock sentiment wordpairs to classify stock news into the positive, negative and neutral news. A wordpairs is a tuple of size 3, consists of a keyword, a polarity word and a sentiment. A keyword usually is a noun or a verb indicating characteristics of stock or business, such as *"profit, recommend, income, price, growth …"* and so on. A polarity word is a verb or an

adjective or an adverb for the keyword above, such as *"good profit, recommend buy, steady income …"* and so on. In this paper, we have only three sentiments: positive (1) if the news has + sign, negative (-1) if the news has – sign, and neutral (0) if the news has 0 sign. The examples of wordpairs are shown in Table 4.

Keyword	Polarity word	Sentiment
กำไร kum-rai 'profit'	ดี dee 'good'	+
คาด kard 'forecast'	กำไร kum-rai 'profit'	+
ปัจจัย pud-jai 'factor'	หนุน nun 'support'	+
แนะนำ nae-num 'recommend'	ถือ thux 'hold'	0
รายได้ rai-dai 'income'	ทรงตัว song-tua 'steady'	0
ราคา ra-ka 'price'	พักตัว puk-tua 'dormancy'	0
ผลกระทบ pon-kra-tob 'effect'	เชิงลบ chung-lop 'negative'	-
เศรษฐกิจ set-ta-kit 'economy'	ชะลอ cha-loor 'slow down'	-
ราคาหุ้น ra-ka-hoon 'stock price'	เสี่ยง seiyng 'risky'	-

Table 4: An example of wordpairs with a pattern {keyword, polarity word, sentiment}

We obtain the first set of wordpairs by hand – called *manual extraction* set (ME). Next, we add new wordpairs into the first set by duplicating the same keyword augmented with an opposite and a neutral polarity words. The second set is called manual wordpairs addition set or *manual addition* (MA), for short. For example, a keyword and polarity word ราคาหุ้น,ขึ้น ra-ka-hoon-,-khun 'stock price, up', the negation polarity word ราคาหุ้น,ลง ra-ka-hoon-,-lng 'stock price, down' and a neutral polarity word ราคาหุ้น,คงที่ ra-ka-hoon-,-kong-thi 'stock price, unchanged'. The MA set has only the positive and negative sense and does not have the neutral sense of 'stock price'. Therefore, this sense is added in the second set. Other examples are shown in Table 5.

The third set of wordpairs is automatically generated from the second set. Wordpairs with partial common keywords are assigned with the same polarity words and signs. For instance, the keywords ราคาหุ้น ra-ka-hoon 'stock price' and ราคา ra-ka 'price' have a common word 'ราคา – stock price'; hence, they will share the same set of polarity words and sentiments.

Keyword	Polarity word	Sentiment
กำไร kum-rai 'profit'	ทรงตัว song-tua 'settled'	0
	ขึ้น khun 'up'	+
	ลง lng 'down'	-
การลงทุน karn-lng-thun 'investment'	ฟื้นตัว fun-taw 'recover'	+
	ซบเซา sob-sea 'stagnant'	-
	คงที่ khong-thi 'stable'	0
แนะนำ nae-num 'recommend'	ขาย khai 'sell'	-
	ซื้อ sux 'buy'	+
	ถือ thux 'hold'	0

Table 5: The manual wordpairs addition with opposite and neutral polarity words

Keyword	Polarity word	Existing wordpairs	New wordpairs
ราคาหุ้น ra-ka-hoon 'stock price'	ขึ้น khun 'up'	ราคาหุ้น, ขึ้น ra-ka-hoon-,-khun 'stock price, up'	ราคา, ขึ้น ra-ka-,-khun 'price, up'
ราคาหุ้น ra-ka-hoon 'stock, price'	บวก bawk 'positive'	ราคาหุ้น, บวก ra-ka-hoon-,-bawk 'stock price, positive'	ราคา, บวก ra-ka-,-bawk 'price, postive'
ราคา ra-ka 'price'	ปรับลด prub-rod 'diluted'	ราคา, ปรับลด ra-ka-,-prub-rod 'price, diluted'	ราคาหุ้น, ปรับลด ra-ka-hoon-,-prub-rod 'stock price, diluted'

Table 6: The automate wordpairs combination

Examples of these crossovers are shown in Table 6. We called the third set automate wordpairs combination or *automate combination*

(AC), for short. The numbers of wordpairs for the three sets are 133, 277 and 331 respectively.

4 Experimental Design and Result

We hypothesized that wordpairs affecting the stock news sentiment can be located near the stock symbol. Hence, we design experiments, using the stock symbol as a center, with different window sizes varying from 20, 40, 60 and 80 letters. Figure 1 demonstrates effects of varying window sizes when extracting wordpairs features.

เรายังคงแนะนำซื้อ MINT CENTEL และ ERW

←←10→MINT←10→
←←←←←←20→CENTEL←20→→→
←←←←←←←←←←←←←←30→ERW←30→→→

range 20 extracts:
[แนะนำ][ซื้อ][MINT] → '[recommend][buy][MINT]'

range 40 extracts:
[แนะนำ][ซื้อ][MINT] → '[recommend][buy][MINT]'
[แนะนำ][ซื้อ][CENTEL] →
'[recommend][buy][CENTEL]'

range 60 extracts:
[แนะนำ][ซื้อ][MINT] → '[recommend][buy][MINT]'
[แนะนำ][ซื้อ][CENTEL] →
'[recommend][buy][CENTEL]'
[แนะนำ][ซื้อ][ERW] → '[recommend][buy][ERW]'

Figure 1: Effects of varying window sizes for extracting wordpairs features

We found that the optimal window size for extracting wordpairs features is 60, with average 1-3 wordpairs for each stock symbol, as shown in Figure 2. We also found that 80-letter window size sometimes extracts irrelevant wordpairs, such as wordpairs of the next symbol. For example, the stock news on Figure 2 has a stock symbol (1) and three wordpairs. Notice that a keyword and a polarity word of the second wordpairs (3) (forecast and profit) are not immediately adjacent to each other.

(1)MAKRO กำไร(2)ดี(2)กว่าคาด(3) คงกำไร(3)ปีนี้ / คงคำแนะนำ(4) ถือ(4)
(1)MAKRO → symbol
(2)กำไร, ดี kum-rai-dee → wordpairs 'profit, good'
(3)คาด, กำไร kard-kum-rai → wordpairs 'forecast profit'
(4)แนะนำ, ถือ nae-num-thux → wordpairs 'recommend, hold'

Figure 2: An example of wordpair features for a stock

In short, there are 6 combinations of {symbol (S), keyword (K), polarity (P)} as shown in Table 7. In the training set, the most frequent patterns found is pattern#3 (K-P-S) with 335, 387, and 387 occurrences when extracted by ME, MA and AC set, respectively. The second most frequent is pattern#5 (P-K-S) with 215, 254, and 257 occurrences when extracted by ME, MA and AC set, respectively. The least frequent pattern found is pattern#4 with 47, 55 and 57 occurrences.

The most and the second most frequent patterns in the testing set found are similar to the training set with 599, 674, and 675 occurrences in pattern#3 and 577, 669, 669 occurrences in pattern#5.

There are two main experiments. The first experiment is designed to examine the effects of the three sets of wordpairs, manual extraction (ME), manual addition (MA) and automate combination (AC), as described in section 3 in training and testing set.

The second experiment is designed to examine the effects of S-K-P patterns on training and testing set. We use an open source software 'Weka' (Weka, 2013) to build classification model using decision tree and support vector machine.

For the first experiment, the results of decision tree and SVM classification models using ME, MA and AC wordpairs sets as features are shown in Table 8. We found that, for the training set, there are no significant differences for all three sets of wordpairs features.

Pattern#	{Symbol, keyword, polarity} combinations
1: SKP	[RATCH][แนะนำ] [ถือ] ราคาเป้าหมาย 64 บาท… [Symbol] [Keyword][Polarity] 'RATCH hold with a target price of 64 Thai baht…'
2: SPK	[SAMTEL] SAT [กำไร] ตาม [คาด]… [Symbol] [Polarity][Keyword] 'SAMTEL and SAT profit as expected…'
3: KSP	รายงานกลุ่ม Small Cap [คาด] [UNIQ] และ TRC มีโอกาสทำ [กำไร]… [Keyword][Symbol] [Polarity] 'small cap segment reported UNIQ and TRC are expected profitable opportunities…'
4: KPS	คงคำ [แนะนำ] [ซื้อ] [TP] 24 บาท… [Keyword][Polarity][Symbol] 'maintain buy with TP 24 Thai baht…'
5: PSK	เรา[ลด] [ราคา]เป้าหมาย และคำแนะนำ [SIM] และ SAMTEL ลงเหลือถือจากซื้อ… [Polarity] [Keyword] [Symbol] 'we lower our target price and recommendation SIM and SAMTEL to hold from buy'
6: PKS	เดือน สค. พบว่า[ปรับสูงขึ้น] 0.5% และ +4.3% โดย [BBL] BAY TMB KBANK SCB รายงาน[สินเชื่อ]เติบโต [Polarity] [Symbol] [Keyword] 'In august, showed a rise of 0.5% and +4.3% BBL BAY TMB KBANK SCB loan growth…'

Table 7: The types of structure matching with the real Thai stock news

The first set – ME, gives a slightly better precision of 0.741 for decision tree model comparing to MA and AC with the precision of 0.722 and 0.721. The same is hold for SVM model, where ME give a slightly better precision result of 0.723 comparing to 0.712 for MA and AC. Every model has the same recall and F-measure of approximately 0.75 and 0.66 respectively. This is not surprising because wordpairs in the first set – ME – are extracted from the training set. Therefore, the first set of wordpairs, when use to classify the training set, should be the most accurate features.

Set	Decision tree		SVM	
ME	Precision	0.741	Precision	0.723
	Recall	0.758	Recall	0.755
	F-Measure	0.669	F-Measure	0.663
MA	Precision	0.722	Precision	0.712
	Recall	0.757	Recall	0.755
	F-Measure	0.669	F-Measure	0.665
AC	Precision	0.721	Precision	0.712
	Recall	0.757	Recall	0.754
	F-Measure	0.669	F-Measure	0.664

Table 8: The result of decision tree and SVM classification of each pattern

The resulting decision tree is partially shown in Figure 3.

Since the precision results of decision tree are better than SVM, we use these training models to classify the testing set. We found that, from 3,489 stock news (# of rows) in the testing set, ME, MA, and AC models predict the same outcome for 3,457 rows or 99.08%; predict the same outcomes two out of three times for 32 rows or 0.91%. There is no totally different outcome prediction – all three models predict different outcomes 0%.

We assume that if three classification models predict the same outcome, then there is no need to verify more. Now, we examine the 32 rows (same two out of three) by comparing the classification models outcomes with solutions provided by a human. The majority outcomes (two same outcomes) is correct for 12 times; while the minority outcome (one out of three) is correct for 16 times. The leftover 4 rows are those outcomes not matched to the solutions. These will be analyzed in section 5.

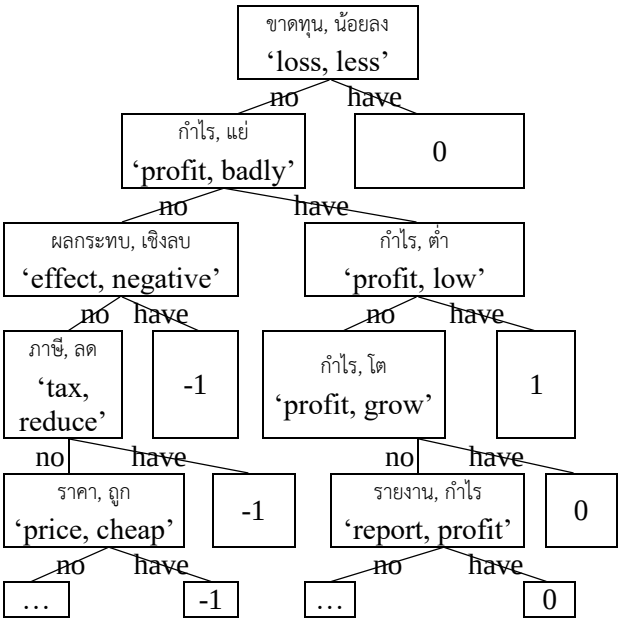

Figure 3: Partial decision tree model built from ME wordpairs features

Wordpair Features		ME			MA			AC		
		Precision	Recall	F-Measure	Precision	Recall	F-Measure	Precision	Recall	F-Measure
Decision tree	P#1:SKP	0.68	0.69	0.66	0.77	0.80	0.77	0.77	0.80	0.77
	P#2:SPK	0.93	0.92	0.91	0.90	0.91	0.90	0.90	0.91	0.90
	P#3:KSP	0.63	0.64	0.62	0.58	0.61	0.58	0.58	0.61	0.58
	P#4:KPS	0.75	0.76	0.74	0.74	0.75	0.72	0.74	0.75	0.72
	P#5:PSK	0.65	0.74	0.69	0.63	0.69	0.66	0.68	0.71	0.69
	P#6PKS	0.75	0.76	0.75	0.79	0.80	0.79	0.76	0.77	0.76
	average	**0.73**	**0.75**	**0.73**	**0.74**	**0.76**	**0.74**	**0.74**	**0.76**	**0.74**
SVM	P#1:SKP	0.70	0.71	0.70	0.81	0.83	0.81	0.79	0.81	0.80
	P#2:SPK	0.96	0.96	0.96	0.95	0.95	0.95	0.95	0.95	0.95
	P#3:KSP	0.70	0.71	0.70	0.69	0.70	0.70	0.69	0.70	0.69
	P#4:KPS	0.70	0.71	0.70	0.71	0.73	0.71	0.71	0.73	0.72
	P#5:PSK	0.77	0.78	0.77	0.72	0.72	0.72	0.75	0.77	0.76
	P#6PKS	0.69	0.71	0.69	0.74	0.76	0.75	0.73	0.75	0.74
	average	**0.75**	**0.76**	**0.75**	**0.77**	**0.78**	**0.77**	**0.77**	**0.79**	**0.78**

Table 9: The result of decision tree and SVM classification of each pattern

For the second experiment, the result is shown in Table 9. Overall, the SVM models give a slightly better results than the decision tree models for all three wordpairs sets. Comparing among the SVM models, the first set – ME, gives the least average recall of 0.76. The second set – MA and the last set – AC give the second best average result of 0.78 and the best average result of 0.79, respectively. However, the result of pattern#2 (S-P-K) obtains very high results above 0.90 for all models. We will discuss insights for this pattern in the next section. The second best result is pattern#1 (S-K-P). It obtains recall of 0.83 and 0.81 when uses with MA and AC set respectively. The other patterns give the similar average results of 0.7 in SVM models.

5 Error Analysis and Discussion

In the first experiment, there are 4 stock news where no machine classification matches the human solutions. The investigation found that these 4 news appear during adjacent business days and have exact same text messages as shown in Figure 4.

กลุ่มรับเหมา… คาดกำไรในไตรมาส 2/58 เติบโตดีอย่าง PTTGC koom-rab-mao-…-kard-kum-rai-nai-tri-mart-song-/-ha-sib-pad-teib-to-dee-yang-pttgc 'Contractor group… Earning expected in the quarter 2/58 growth as PTTGC'	
The wordpairs should be extracted.	เติบโต, ดี teib-to-dee 'growth, good'
The wordpairs is in the ME, MA, and AC sets.	การเติบโต, ดี karn-teib-to-dee 'growth, good'

Figure 4: Error analysis for stock news classification

From the figure 4, the wordpairs เติบโต, ดี teib-to-dee 'growth, good' and การเติบโต, ดี karn-teib-to-dee 'growth, good' have the same meaning. However, in all three wordpair sets, there is no เติบโต, ดี teib-to-dee 'growth, good' as a wordpair feature. Hence, the keyword is not extracted as a feature. The type of เติบโต tieb-to 'growth' is a verb, but การเติบโต karn-tieb-to 'growth' is a noun. In Thai language, the addition prefix of the word การ karn or ความ kwarm will change a type of a word from a verb to a noun. This investigation suggests that this factor should be considered in order to construct a better set of wordpair features for Thai language in the future.

In the second experiment, pattern#2 (S-P-K) has the highest precision, recall and f-measure. An investigation of the training set for pattern#2 (S-P-K) found that the set has only two sentiments: positive (1) and negative (-1). The training set contains no neutral sentiment. Therefore, there will be only two classes of classification results instead of three. The results suggest that the stock news sentiments classification for this pattern – (S-P-K)

can be performed with more accuracy than other patterns because it has only two polarities instead of three polarities.

6 Conclusion and Future Work

This paper proposes to classify stock news into three classes: positive, negative and neutral using only text in the news called wordpairs. Three sets of wordpairs are constructed. The first set is manually extracted from 1,381 stock news. It contains 133 wordpairs. The second set is manually added with opposite and neutral polarities and contains 277 wordpairs. The third set is automatically generated using partial keyword combined with existing polarities and contains 331 wordpairs.

Two experiments are conducted to test the effects of three wordpair sets. The result found no significant differences in the training set but found slightly improvement, for the second the third wordpair sets, when they are applied to unseen stock news (a testing set) from other brokers. Moreover, we found six combination patterns of a stock symbol, a keyword and a polarity (S-K-P) in stock news. The result from the second experiment shows that some pattern (S-P-K) has only two polarities, instead of three and therefore achieved the highest correct classification results.

For the future work, we will resolve the problem found and discussed in the error analysis section by consider adding and deleting a keyword's prefix.

References

Apinan Chattupan and Ponrudee Netisopakul. 2014. Stock sentiment analysis model using data mining (In Thai). In Knowledge and Smart Technology (KST), 2014. Proceeding of 6th National Conference on. Chonburi, Thailand.

Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll and Manfred Stede. 2011. Lexicon-based methods for sentiment analysis. Computational Linguistics, 37(2):267-307.

Marc-André Mittermayer. 2004. Forecasting intraday stock price trends with text mining techniques. In System Sciences, 2004. Proceeding of the 37th Annual Hawaii International Conference on. IEEE.

Nattadaporn Lertcheva and Wirote Aroonmanakun. 2009. A linguistic study of product names in Thai economic news. In Natural Language Processing, 2009. SNLP'09. Eight International Symposium on, 26-29. IEEE.

Nattapong Tongtep and Thanaruk Theeramunkong. 2010. Pattern-based extraction of named entities in thai news documents. Thammasat International Journal of Science and Technology, 15(1):70-81.

Phaisarn Sutheebanjard and Wichian Premchaiswadi. 2010. Disambiguation of Thai personal name from online news articles. In Computer Engineering and Technology (ICCET), 2010 2nd International Conference on, 3:302-306. IEEE.

Rathawut Lertsuksakda, Kitsuchart Pasupa and Ponrudee Netisopakul. 2015. Sentiment analysis of Thai children stories on support vector machine. In Artificial Life and Robotics (AROB), 2015. Proceeding of the Twentieth International Symposium on. Beppu, Japan.

Rathawut Lertsuksakda, Ponrudee Netisopakul and Kitsuchart Pasupa. 2014. Thai sentiment terms construction using the Hourglass of Emotions. In Knowledge and Smart Technology (KST), 2014 6th International Conference on, 46-50. IEEE.

Robert P. Schumaker and Hsinchun Chen. 2009. Textual analysis of stock market prediction using breaking financial news: The AZFin text system. ACM Transactions on Information System (TOIS), 27(2).

Bualuang Securities. Retrieved February 15, 2015. http://www.bualuang.co.th/th/index.php

Stock News Online. Retrieved May 15, 2015. http://www.kaohoon.com/online/content/category/13/ภาวะเศรษฐกิจและตลาดหุ้นในประเทศ

Weka 3.7.1. Retrieved October 1, 2013. http://www.cs.waikato.ac.nz/ml/weka