ABSTRACTIVE THAI OPINION SUMMARIZATION

Orawan Chaowalit

A Dissertation Submitted in Partial

Fulfillment of the Requirements for the Degree of

Doctor of Philosophy (Computer Science)

School of Applied Statistics

National Institute of Development Administration

2013

ABSTRACTIVE THAI OPINION SUMMARIZATION

Orawan Chaowalit

School of Applied Statistics

Assistant Professor ....Major Advisor

(Ohm Sornil, Ph.D.)

The Examining Committee Approved This Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy (Computer Science).

Associate Professor ....Committee Chairperson

(Surapong Auwatamongkol, Ph.D.)

Assistant Professor..Committee

(Ohm Sornil, Ph.D.)

Associate Professor..Committee

(Pipat Hiranvanichakorn, D.E.)

Assistant Professor..Committee

(Rawiwan Tenissara, Ph.D.)

Instructor..Dean

(Siwiga Dusadenoad, Ph.D.)

April 2014

ABSTRACT

Title of Dissertation	Abstractive Thai Opinion Summarization
Author	Miss Orawan Chaowalit
Degree	Doctor of Philosophy (Computer Science)
Year	2013

With the advancement of the Internet technology, customers can easily share opinions about services and products in forms of reviews. There can be large amount of reviews for popular products. Manually summarizing those reviews for important issues is a daunting task. Automatic opinion summarization is a solution to the problem. The task is more complicated for reviews written in Thai. Thai words are written continuously without space, and there is no symbol to identify the end of a sentence. Many reviews are written informally, thus accurate word identification and linguistic annotation cannot be relied upon. Text summarization can be classified into two categories, which are extractive and abstractive summarization. In the extractive method, the summary is a set of actual sentences or phrases extracted from the reviews; on the other hand, abstractive summarization does not output original sentences from the reviews, but generates new sentences or phrases into a summary. The abstractive summarization approach is more difficult and thus less popular than the extractive approach. This research proposes a novel technique to generate abstractive summaries of customer reviews written in Thai. The proposed technique, which consists of local and global models, is evaluated by using actual reviews of fifty products, randomly selected from a popular cosmetic website. The results show that the local model outperforms the global model and the two baseline methods, both quantitatively and qualitatively.

ACKNOWLEDGEMENTS

I would like to express sincere thanks to my major advisor, Assistant Professor Dr. Ohm Sornil, for his valuable advice, encouragement and guidance in making this dissertation and successful one, I also wish to extend thanks and appreciation to all of the committee members, Associate Professor Dr.Surapong Auwaanamongkol, Associate Professor Dr. Pipat Hiranvanichakorn and Assistant Professor Dr.Rawiwan Tenissara, for their thoughtful comments and suggestions. I am indebted to all lecturers and staffs at the School of Applied Statistics, National Institute of Development Administration.

Gratitude is also dedicated to Silpakorn University, for providing me a scholarship to study at the Nation Institute of Development Administration (NIDA). I would like to my former teachers, my friends, and my colleagues for help and encourage me.

Special thanks are due to my parents for their support and care.

Orawan Chaowalit

May 2014

TABLE OF CONTENTS

LIST OF TABLES

Tables **Page**

LIST OF FIGURES

Figures **Page**

CHAPTER 1

INTRODUCTION

Nowadays consumers can buy products and services from vendors around the world; therefore, companies need to secure and expand their market shares. Customer satisfaction is critical, and to maintain customer base, continuous development and improvement on their products is crucial. On the customers' side, they want information from others to decide whether the products are good or suitable for them or not. In the past, customers used to ask their friends and families, who had used the product, for their opinions. Companies hired market researchers to survey their customer satisfactions. With the advent of the Internet, product reviews can be done through web boards, web blogs, or the companies' third-party websites. This makes it easier and more convenient for companies to have necessary information to improve their products, and customers also have aids for their buying decisions.

The number of reviews has increased to over a hundred reviews per product or more, if the product is popular, or if the websites are well known and trusted by customers. Customer reviews are mostly unstructured, and usually use natural language texts. Some are long, some are short, some are grammatically correct sentences, and some are short phrases. Reviews are usually about details and properties of the products and are sometimes redundant in contents to indicate that those properties are crucial to customers. Due to the amount of reviews, it is difficult for readers to summarize main ideas of those opinions. An automatic opinion text summarization is a solution to this problem by summarizing core ideas of the entire reviews.

1.1 Statement of the Problem

Large amount of online opinions make it difficult for the readers to digest the information precisely enough to consider. In this task, we will be discussing about readers who are users that want to find some liquid foundation. They will select the brand, and each brand has varieties of customer reviews. Thus, each user has significant points of interest, or specific need, e.g., a user with dry skin may write that the product is good; on the contrary, a user with oily skin may write bad review about the product. That is a challenge to users to understand and make decision. Therefore, the opinion summary process technique help readers to do research, process the information and summarize all interesting and important opinions.

1.2 Objectives of the Dissertation

The objective of this dissertation is to develop an alternative methodology to extract important content from the Thai reviews. The reviews are highly redundant and have no specific entities. The methods of summarizing the reviews are to select and rewrite a subset of the original sentences from the reviews, capture their main points from redundant opinion and rewrite them into abstraction summarization. This technique can be applied and used with verities of products or services that have high duplicate data.

CHAPTER 2

BACKGROUND AND RELATED WORK

In this thesis, we propose the technique of abstracting Thai opinion summarization in high duplicate context. The method comprises of text extraction, text segmentation and text abstraction. Text summarization will be explained first, then characteristic of opinion text and opinion text summarization respectively. Related work with this thesis method will be discussed last.

2.1 Text Summarization

Text summarization, in general, can be classified into two types: extractive and abstractive approach. In an extractive method (Sornil and Gree-ut, 2006:1-6; Carenini and Cheung, 2008: 33-41) the summary is a set of actual text segments extracted from the reviews. Most research in this category focuses on texts relevant to the title or a topic of interest (Sornil and Gree-ut, 2006:1-6), and the output is a set of direct sentences in the top ranks taken from the original review texts, according to some criteria. In addition, the technique is to extract relevant sentences and use information retrieval (IR) technique; such as PageRank (Brin and Page, 1998: 107-117), HITS (Kleinberg, 1999: 604-632) and combine with term frequency inverse document frequency (tfidf) (Salton, Singhal, Buckley and Mitra, 1999: 53-65; Banko, Mittal, Kantrowitz, and Goldstein, 1999) in such a position of sentence, first sentence in paragraph and similarity between sentence and title (Ishikawa, ANDO, Doi, and Okumura, 2002; Shen et al., 2004: 242-249). Another technique uses part of speech (POS) (Lin and Eduard, 2003: 1-8) to find some words that have the same meaning but different in their part of speech; for example, noun and verb. Abstractive summarization (Ganesan, Zhai, and Han, 2010: 340-348; Carenini and Cheung, 2008: 33-41; Lloret, Romá-Ferri, and Palomar, 2013: 164-175; Luhn, 1958: 159-165) does not output sentences from original reviews, but generates new sentences or phrases as a summary. This approach is more difficult and thus less popular than the extractive approach.

2.2 Abstractive Text Summarization

As mentioned earlier, abstract text summarization uses some of the techniques from text summarization. Thus, primary technique in traditional summarization was used, such as select key phrase or key word, or use method selection from traditional text summary. There are two approaches in abstracting similar multi-sentence compression. The first one is selecting the significant sentence, and then selects another sentence that is significantly different. Second, make up a cluster, and then select one sentence in that cluster to use as a summary; however, irrelevant sentence may be selected. Whereas Filippova (2010: 322-330) presents a multi-sentence compression method using directed graph with the shortest path algorithm based on term frequencies, and a simple grammatical checking process is included for English and Spanish. All data is kept from start to end with node, which contains words and edges that represents the link between word A and word B. After the first sentence was added to the graph, another sentence will be added consequentially. The word in the new sentence will be added to the node if it did not exist in the graph. The number of words will be counted and put into numbers. Figure 2.1 shows an example of the sentence in the node graph.

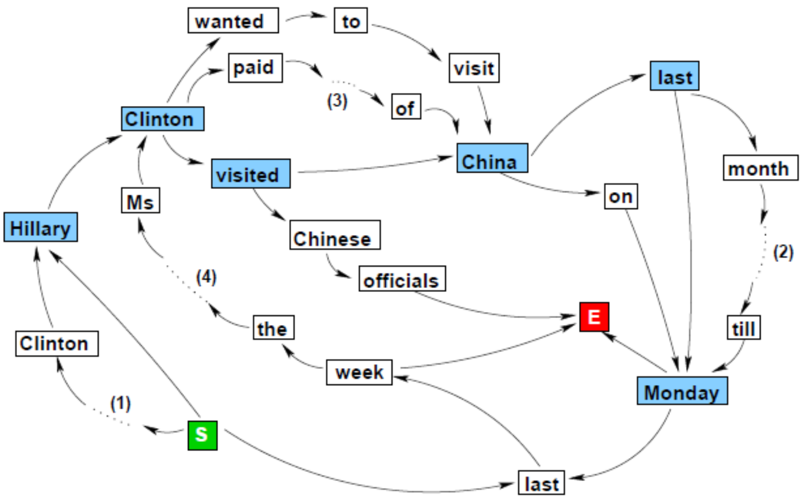

Figure 2.1 Word Graph Generated from Sentence and Possible Compression Path

Source: Filippova, 2010: 324.

There is part of speech (noun, verb, adjective, and adverb) in the graph that keeps the position of each word in the sentence. The graph traverses for the shortest path algorithm, based on term frequencies. They used weight of node according to the following equation

$$w(e_{i,j}) = \frac{(freq(i) + freq(j))}{freq(e_{i,j})}$$

where

$w(e_{i,j})$ is the edge weight between word *i* and word *j*

$freq_{(i)}$ is the frequency of $word_{(i)}$ in document.

$freq_{(ei,j)}$ is the frequency of $word_{(i,j)}$ in document in the same time.

During traversing the graph, the grammar was checked. In the end, the path with the highest score and the most correct grammar was selected. This process has been experimented with English and Spanish.

COMPENDIUM (Lloret, Romá-Ferri, and Palomar, 2011: 61-66, 2013: 164-175) was proposed for text summarization to generate abstracts of biomedical research papers. There were two stages in the approach; stage (i) extraction summarization system uses basic linguistic for text analysis and tokenization, and sentence segmentation. After that, redundancy word or repeated information was discovered. A graph was created from the word node that was shown in Figure 2.2. First node in each sentence was initialized to the first node graph and each node word was mapped to the same node.

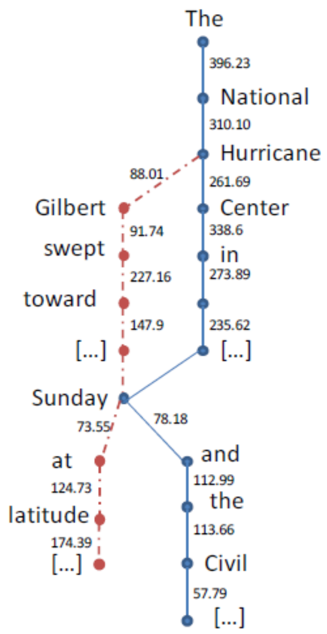

Figure 2.2 Example Graph Word Node Crated from Sentence

Source: Lloret, 2011: 62.

Next, the sum of the relevant topic was identified with term frequency. Thereafter, all sentences were computed into a s core, using the shortest path algorithm, e.g., Dijkstra's algorithm and weight of word can be computed into the following equation

$$W(e_{i,i+1}) = \frac{1}{(FreqRel_{(i,i+1)}) * (PR_i + PR_{(i+1)})}$$

where

$W(e_{i,i+1})$ is the edge weight of the edge connecting $word_{(i)}$ and $word_{(i+1)}$.

$FreqRel_{(i,i+1)}$ is the frequency link between $word_{(i)}$ and $word(_{i+1})$

$PR_{(i)}$ is PageRank $word_{(}v_{i)}$ and PageRank $word_{(i+1)}$

Finally, the summary was generated from high-ranking score. On stage (ii), the graph generates with word similar approach, which can be explained as Filippova. After that, the sentence was generated and incorrect grammar sentences were rejected. Last, weight sentence was computed and the best one is selected to be the summary. In other words, a group of sentences are summarized into one sentence.

2.3 Opinion Texts

An opinion is a view or judgment formed about objects or features on something. Some opinion texts may be redundant in meaning. Opinion text has a unique characteristic that is different from standard text. The standard or classic text usually has a title; either the first sentence or the last sentence regularly indicates the importance. The opinion text usually has a polarity words, e.g., good or bad to express about the entities of the product or service. Many of texts come from variety of users or customers with vast difference in personalities; consequently, some users have other entities for the product. Some research had tried to solve this problem by implementing the technique called sentiment analysis.

However, in this dissertation, we are focusing on opinion summarization. Figure 2.3 shows the example of user reviews in Thai and Figure 2.4 shows the example of user reviews in English. In these two pictures, we can see that in the context of the opinion, reviewers usually write duplicate text; for example, “ระหว่างวันไม่ดรอป”, “สีไม่ดรอประหว่างวัน”. In Figure 2.3, “cakey” means thickness when apply the foundation on face. In Figure 2.4, from the example, we can see that the opinion text is written with duplicate text from many reviewers.

Figure 2.3 Example of reviews in Thai about cosmetic products

Source: jeban, 2014.

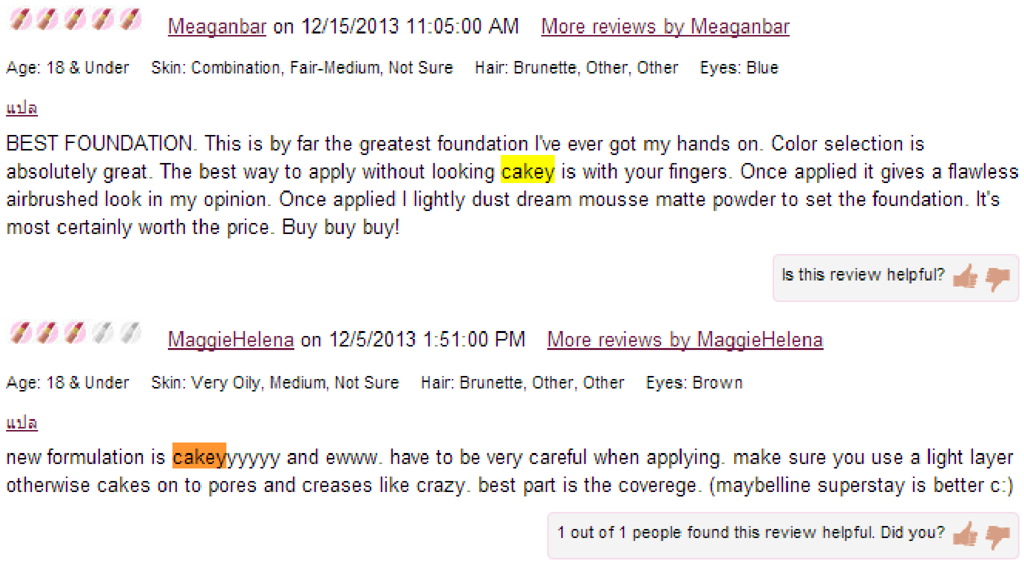

Figure 2.4 Example of reviews in English about cosmetic products

Source: MakeupAlley, 2014.

For opinion summarization, there are large numbers of reviews written by customers, and they are redundant in nature. Most summarization research generates phrases or short sentences that can convey information. Ganesan, Zhai and Han (2010: 340-348) used an unsupervised method and a graph structure, created from words in the reviews, part of speech, and locations of terms in the original sentences, to generate sentences, according to a topic of interest, such as iPhone battery life. A graph is traversed to generate a summary whose grammar is checked against four predefined templates. Acceptable sentences are then scored and ranked based on term frequencies. Figure 2.5 shows an example of graph and generated result. In this case, the result is “The iPhone is a great device and is worth the price”.

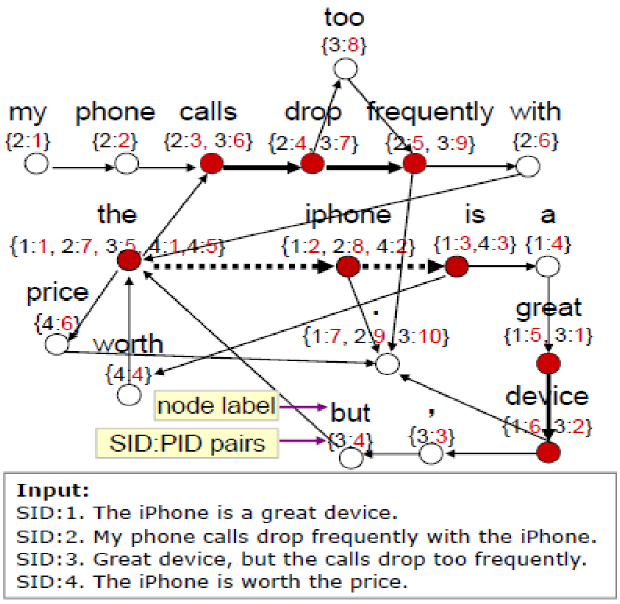

Figure 2.5 Sample Opinosis Graph (Thick edges indicate salient paths)

Source: Ganesan, Zhai, and Han, 2010: 342.

Liu, An, and Song (2011: 2026-2031) had proposed, "Chinese Multi-document Summarization Based on O pinion Similarity". This technique extracts opinion and opinion similarity. First, this method extracts opinion from original customer reviews and the sentence was calculated and scored from redundant attribute information, then the highest score sentence was selected as the summary opinion. Finally, the summary was generated with significant attribute or entities and each sentiment word.

Micropinion (Ganesan, Zhai and Viegas, 2012: 869-878) generates understandable short phrases of two to seven words long depending on device display by using a publicly available n-gram model and a d epth first search to concatenate seed bigrams while sentences are structurally examined. By using input from similar sentence and having the same features such as "battery life being excellent", scores for sentences are calculated from probability of terms to occur together and term readability, without the use of linguistic annotations.

Most opinion summarization works are proposed for Western languages; however, in Thai, words are written continuously without space. Identifying word boundary is shown to be a difficult and inaccurate task, thus word identification, post tagging, and grammar checking cannot be relied on in the technique. There has not been research on methods to abstractive summarize Thai opinion texts. This research proposes a technique to solve that problem. The technique begins with an extraction of important text segments representing opinions of users using a graph ranking algorithm. Bigrams are generated from the segments, and relationships among them are used to create a graph. Through a traversal, a set of text phrases are created and ranked according to word importance, collocations among words, and structures of review texts. Highly similar phrases are grouped, and the top-ranked phrase in each group is included in the summary. Our technique is fully unsupervised, domain independent, and not using grammar or linguistic annotation; nevertheless, relying on important text segments extracted, redundancies and writing structures in the review texts, written in Thai language.

CHAPTER 3

METHODOLOGY

This research proposes a technique to generate a rewritten summary of unstructured opinions from online customer reviews written in Thai. The proposed technique consists of 3 main processes: (1) segment extraction, (2) candidate phrase construction, and (3) summary generation. The segment extraction process selects important text segments that convey meanings of the reviews by creating a weighted segment graph and ranking those segments by using the Hopfield network algorithm. High-ranked segments are considered important and thus extracted. Bigrams from the selected segments with strong relationships with others are used to create a graph, which is traversed to construct a set of candidate phrases. These phrases are scored and ordered considering word importance, word collocations, and writing structures. Similar phrases are grouped together, and the top-ranked segment from each group is included in the summary. The outline of the system architecture is shown in Figure 3.1.

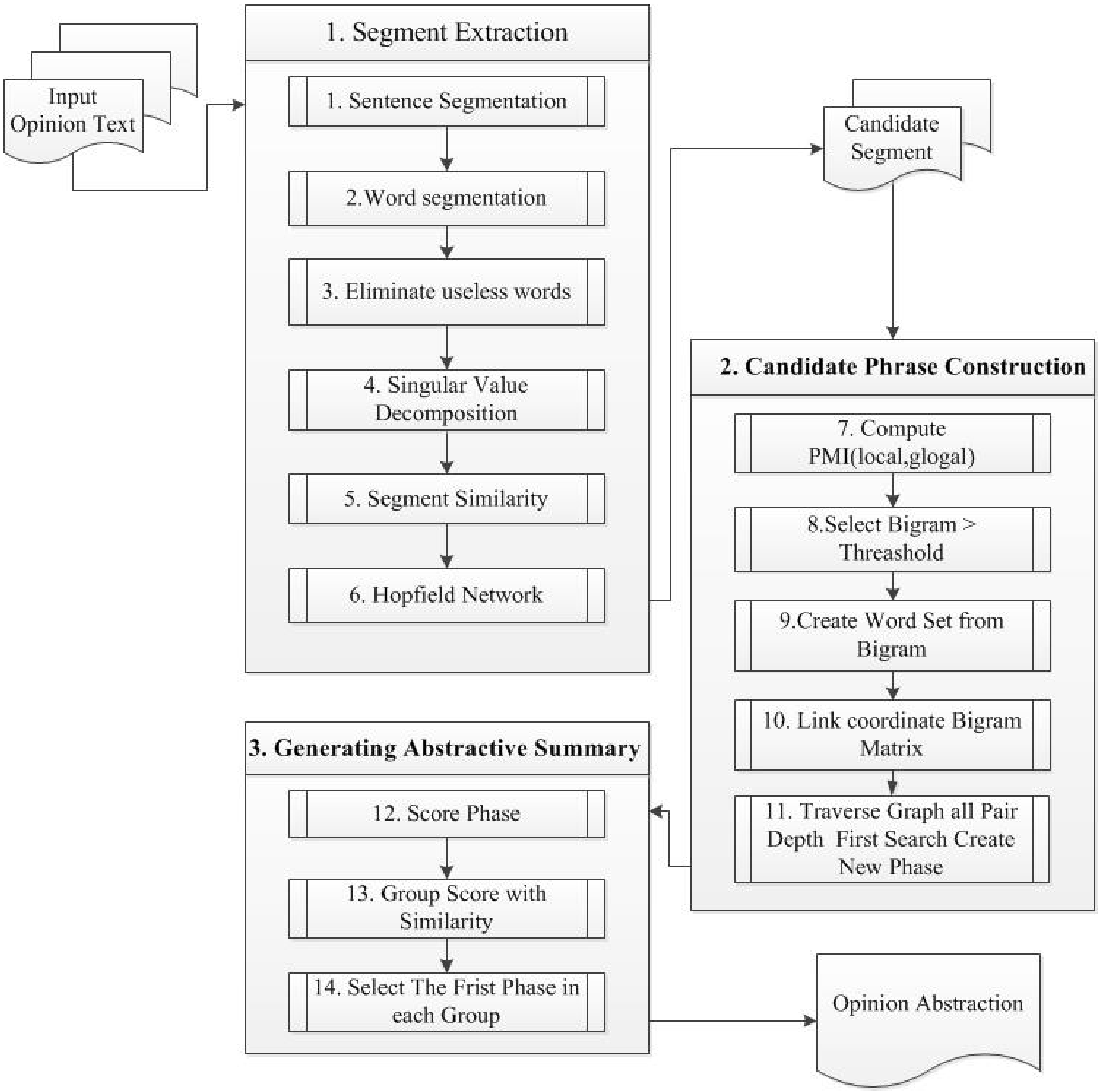

Figure 3.1 Outline of System Architecture

3.1 Segment Extraction

The purpose of this step is to extract important segments captured from a set of reviews from customer opinions.

3.1.1 Thai Word Identification

In Thai, there is no symbol to identify the end of a s entence. Reviews are written in free forms. Though understandable by readers, they are generally not complete sentences or grammatically correct sentences. In this research, text segments are character strings from customer reviews, which are separated by special symbols ("?", ".", ", ", ", ", "*","-", or whitespace), as defined in (Sornil and Gree-ut, 2006). In addition, unlike in English word segmentation, in Thai, and in many other Asian languages, it is more complex because the language does not have any explicit word boundary delimiters, such as space to separate between words; for example, "ฉันทานข้าวและขนมอิ่มมากจนทานอะไรไม่ไหวอีกแล้วคืนนี้". Figure 3.2 shows an example of input opinion data text.

น่าอ่านมาก อยากเป็นครู เป็นคุณครู อ่านแล้วช่วยให้ทำงานเป็นขึ้นเยอะเลยเผื่อได้เลื่อนขั้นกะเขาบ้าง เป็นหนังสือที่ดีมากสำหรับผู้ที่ต้องการเลื่อนตำแหน่งในสายอาชีพราชการครู เป็นหนังสือที่พัฒนาสูตรสำเร็จข้าราชการได้ดีมาก ดีมาก ไม่ใช่ครูก็อ่านและไปปรับใช้ได้นะ ไม่เฉพาะคุณครูนะที่ควรอ่าน อ่านได้ทุกอาชีพ น่าสนใจมาก อ่านแล้วรู้สึกว่าข้าราชการน่าทึ่งมาก เลยคะ อยากเป็นจังเลย อ่านแล้วอยากเป็นข้าราชการมากเลย ได้ความรู้ดี .....

Figure 3.2 Example opinion texts**.**

We use the text from Figure 3.2 as input to cut into sentence segments by using simple algorithm. Figure 3.3 shows the output from the process of sentence cutting into segments.

น่าอ่านมาก
อยากเป็นครู
เป็นคุณครู
อ่านแล้วช่วยให้ทำงานเป็นขึ้นเยอะเลยเผื่อได้เลื่อนขั้นกะเขาบ้าง
เป็นหนังสือที่ดีมากสำหรับผู้ที่ต้องการเลื่อนตำแหน่งในสายอาชีพราชการครู
เป็นหนังสือที่พัฒนาสูตรสำเร็จข้าราชการได้ดีมาก
ดีมาก
ไม่ใช่ครูก็อ่านและไปปรับใช้ได้นะ
ไม่เฉพาะคุณครูนะที่ควรอ่าน
อ่านได้ทุกอาชีพ
น่าสนใจมาก
อ่านแล้วรู้สึกว่าข้าราชการน่าทึ่งมากเลย
อยากเป็นจังเลย
อ่านแล้วอยากเป็นข้าราชการมากเลย
ได้ความรู้ดี ......

Figure 3.3 An Example of Sentence Segments.

After that, word segmentation was performed. There are several levels and several roles for Thai characters that may lead to ambiguity in segmenting the words (Bheganan, Nayak and Xu, 2009: 74-85).

In Thai, characters are written without explicit word boundaries. Depending on the contexts, there can be many ways to break a string into words, for instance, "อาจอง" can be segmented as "อา*จอง" or "อาจ*อง", and "นั่งตากลม" can be segmented as "นั่ง*ตา*กลม" or "นั่ง*ตาก*ลม". This complicates the task of identifying word boundaries.

In this research, word segmentation is done by using the Smart Word Analysis for THai (SWATH) software to select the longest matching algorithm (Charoenpornsawat and Sornlertlamvanich, 2001: 231-235). Figure 3.4 shows output after the SWATH software has performed its task.

น่า|อ่าน|มาก|
อยาก|เป็น|ครู|
เป็นคุณ|ครู|
อ่าน|แล้ว|ช่วย|ให้|ทำงาน|เป็น|ขึ้น|เยอะ|เลย|เผื่อ|ได้|เลื่อน|ขั้น|กะ|เขา|บ้าง|
เป็น|หนังสือ|ที่|ดี|มาก|สำหรับ|ผู้|ที่|ต้องการ|เลื่อน|ตำแหน่ง|ใน|สาย|อาชีพ|ราชการ|
ครู|
เป็น|หนังสือ|ที่|พัฒนา|สูตร|สำเร็จ|ข้าราชการ|ได้|ดี|มาก|
ดี|มาก|
ไม่|ใช่|ครู|ก็|อ่าน|และ|ไป|ปรับ|ใช้ได้|นะ|
ไม่|เฉพาะ|คุณครู|นะ|ที่|ควร|อ่าน|
อ่าน|ได้|ทุก|อาชีพ|
น่า|สนใจ|มาก|
อ่าน|แล้ว|รู้สึก|ว่า|ข้าราชการ|น่าทึ่ง|มาก|
เลย|
อยาก|เป็น|จัง|เลย|
อ่าน|แล้ว|อยาก|เป็น|ข้าราชการ|มาก|เลย|
ได้|ความ|รู้ดี| …..

Figure 3.4 Example of Output from SWATH

Then, some words are removed from the output into a set of shortlist of stop words. The shortlist of stop words is used, which include non-meaningful words:
นะค่ะ, นะคะ, อะคะ, อ่ะคะ, อะค่ะ, อ่ะค่ะ, อ่ะ, ค่ะ, คะ, นะครับ, ครับ, ๆ, ๆ, T, ^ ,) , (, ., : , @, -, _

We use only words in the shortlist because Thai characters are different from English. For example, "ก็" which means; too; as well; well; may, can also possibly be used as a particle to emphasize something that the word follows. Thus, in this research, text data is huge and it is beneficial to eliminate some words with confusing meaning; consequently, we only use words from the shortlist.

3.1.2 Word-Segment Matrix Compression

Each text segment S_j is represented as a vector $<f_{1j}, f_{2j}, \ldots, f_{mn}>$, where f_{ij} is the frequency of term i in segment j, m is the total number of terms in the entire reviews, and n is the number of segments in the reviews. A word-segment matrix with n rows and m columns are created, which is shown in Figure 3.5.

w\segment	อ่าน	แล้ว	อยาก	เป็น	ข้าราชการ	ครู
s1	1	1	1	1	1	
s2			1	1	1	
s3	1	1	1	1		1
s4	1	1				

Figure 3.5 An Example of Word-Segment Matrix

The organization of the matrix assumes that all words are independent, which may not generally be true in practice. Also, with a large number of words, further processing is computationally expensive due to high dimensionality. A Singular Value Decomposition (SVD) (Baeza-Yates, Araújo Neto Ribeiro and Ribeiro-Neto, 1999: 44-45) is performed to compress the matrix into a lower dimensional feature space that can uncover hidden relationships among features and segments, and reduce effects of noises in segment characteristics. SVD decomposes matrix A into three components: an orthogonal matrix of singular values, where $r = min\ (m, n)$, and the left and the right singular vectors (i.e., U and V, respectively), as shown in Figure 3.6. By keeping $k < r$ largest values of the singular matrix along with their corresponding columns in U and V, the resulting matrix is a matrix of rank k, which is closest to the original matrix A in the least square sense. With respect to this new space of k dimensions, the attributes are no longer independent from each other.

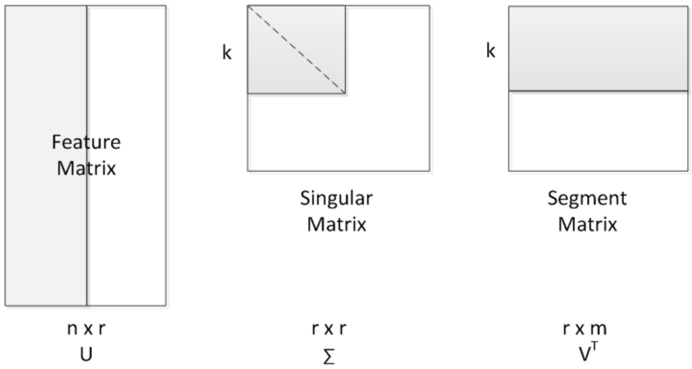

Figure 3.6 Singular Value Decomposition

Definition: Let A can be an m n matrix then the *singular values* of A are defined to be the square roots of the eigenvalues of A^TA. The singular values of A will be denoted by $\sigma_1, \sigma_2, \ldots, \sigma_n$. It is customary to list the singular values in decreasing order so it will be assumed that $\sigma_1 \geq \sigma_2 \geq \quad \geq \sigma_n \geq 0$

Theorem: Let A be any n m matrix. Matrix A can be written in $A = U\Sigma V^T$ where U is an n m orthogonal matrix, V is an n m orthogonal matrix, and Σ is an m n matrix whose first r diagonal entries are the non-zero singular values $\sigma_1, \sigma_2, \ldots, \sigma_r$ of A and all other entries are zero.(whose other entities are all zero) . The expression $U\Sigma V^T$ is known as the Singular Value Decomposition. The columns of V are called the *right singular vectors*. The columns of U are called the *left singular vectors*.

$$A_{\mathrm{nm}} = U_{nr}\Sigma_{\mathrm{rr}}V_{rm}^T$$

Where $U^TU = I.V^TV = I$; the columns of U are orthonormal eigenvectors of A^T , the columns of V are orthonormal eigenvectors of AA^T, and S is a diagonal matrix containing the square roots of eigenvalues of U or V in descending order.

In this content, matrix A is word $\times$ document matrix. From original matrix was performed to three matrices that show in Figure 3.7.

$$\begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} =$$

$$\begin{bmatrix} -0.542 & -0.454 & -0.707 \\ -0.542 & -0.454 & 0.707 \\ -0.643 & 0.766 & 0.000 \end{bmatrix} \begin{bmatrix} 2.524 & 0 & 0 \\ 0 & 0.792 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -0.684 & -0.211 & -0.684 \\ -0.180 & 0.789 & -0.180 \\ -0.707 & 0.000 & -0.707 \end{bmatrix}$$

$$\begin{bmatrix} 1.001 & 0.001 & 1.001 \\ 1.001 & 0.001 & 0.001 \\ 1.001 & 1.001 & 1.001 \end{bmatrix} \approx$$

$$\begin{bmatrix} -0.542 & -0.454 \\ -0.542 & -0.454 \\ -0.643 & 0.766 \end{bmatrix} \begin{bmatrix} 2.524 & 0 \\ 0 & 0.792 \end{bmatrix} \begin{bmatrix} -0.684 & -0.211 & -0.684 \\ -0.180 & 0.789 & -0.180 \end{bmatrix}$$

Figure 3.7 Example of Matrix a Decomposes into Three Matrices

The purpose of SVD is to actually reconstruct the original matrix, in order to suppress noise in the original matrix.

3.1.3 Segment Extraction

A segment graph G is constructed from the compressed word-segment matrix where $G = (V, E)$ is a document graph with a set of vertices V and a set of edges (or links) E where V = {S_1, S_2, ...,S_n}; S_i is segment i in the document; and E is a subset of $V \times V$. A segment S_i is defined as a vector $<f_{1i}, f_{2i}, \ldots, f_{ki}>$ where f_{ki} is the value of feature (word) l for segment i, and k is the total number of features (words).

Degree of similarity between segment S_i and segment S_j becomes the edge weight between nodes representing the two segments; it can be calculated as follows:

$$similarity(S_i, S_j) = \frac{\sum_{v=1}^{k} f_{s_{iv}} * f_{s_{jv}}}{\sqrt{\sum_{v=1}^{k} f_{s_{iv}}}\sqrt{\sum_{v=1}^{n} f_{s_{jv}}}}$$

The segment graph is an undirected weighted graph with edges placed between segments with sufficient similarities, as shown in Figure 3.8.

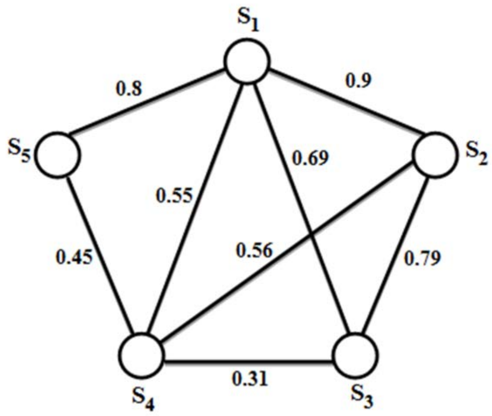

Figure 3.8 Segment Graph

Each segment node will be assigned a significant score, using a graph ranking algorithm, the Hopfield network algorithm (Chen and Ng, 1995: 68-73). The algorithm performs a p arallel relaxation search, in which nodes are activated in parallel, and activation values from different nodes are combined for each individual node. Neighboring nodes are traversed in order until the activation levels of nodes in the network converge. In the context of a segment graph, the graph can be viewed as a network whose nodes are represented by neurons, and edges are represented by synaptic links. The process terminates when there is no significant difference in terms of output between two consecutive iterations. The algorithm can be described as follows:

Initial State: The algorithm is initialized by

$$u_i(0) = 1,\ 0 \leq i \leq n\text{-}1$$

where $u_i(t)$ is the score of node i at iteration t.

Activation and Update State: Output of each node is calculated as follows:

$$u_i(t+1) = sigmoid[net_j] \text{ , } 0 \leq j \leq n\text{-}1$$

where $net_j = \sum_{n=0}^{n-1} w_{ij}\, u_i(t)$ is input through the activation function, and w_{ij} is the weight of the synaptic link between S_i and S_j, and

$$sigmoid[net_j] = \frac{1}{1 + exp[\frac{\theta_j - net_j}{\theta_j}]}$$

where θ_j is a bias and θ_o is an adjustable constant.

Stable State: Repeat the iteration until convergence. The stable state is achieved when sum of the error at every node in the network falls below a given threshold (ε).

$$\sum_{j=1}^{n-1} | u_j(t+1) - u_j(t) | \leq \varepsilon$$

Outputting State: After the network converges, the resulting outputs become the final significance scores of the corresponding segments for extraction.

Once the algorithm terminates, we have a score u_i for every segment i. The scores are sorted in a descending order. Segments with top R segments are selected as the source of abstraction in further steps. The parameter R will be studied in the experiments.

After this step, a segment ranking score is performed and the significant high score is selected for the source to generate a phrase. Figure 3.9 shows an example of the output.

เป็น|หนังสือ|ที่|ดี|มาก
อ่าน|แล้ว|อยาก|เป็น|หนังสือ|ที่|ดี|มาก
น่า|อ่าน|แล้ว|อยาก|เป็น|หนังสือ|ที่|ดี|มาก
อยาก|เป็น|หนังสือ|ที่|ดี|มาก
แล้ว|อยาก|เป็น|หนังสือ|ที่|ดี|มาก
อ่าน|แล้ว|อยาก|เป็น|ข้าราชการ
หนังสือ|ที่|ดี|มาก
น่า|อ่าน|แล้ว|อยาก|เป็น|ข้าราชการ….

Figure 3.9 An Example Output Ranking with Hopfield Network.

3.2 Candidate Phrase Construction

From the set of selected segments, word-based bigrams are extracted and used to create a word graph. The graph is then traversed to generate a candidate phrases set. The word graph is created from bigrams whose preceding word shave strong collocation strengths with the following words. The collocation strength can be calculated using a modified version of Pointwise Mutual Information (PMI) (Damani, 2013: 163-169) which is biased toward bigrams that occur in many segments and represents legitimate word sequences. The collocation strength for a bigram (w_i,w_j) can be calculated as follows:

$$collocation(w_i, w_j) = freq(w_i, w_j) * \log_2\left(\frac{P(w_i, w_j)}{P(w_i) * P(w_j)}\right)$$

where $P(w_i, w_j)$ is the co-occurrence probability of the bigram; $P(w_i)$ and $P(w_j)$are the probabilities of occurrences of w_i and w_j, respectively; and $freq(w_i,w_j)$ is the co-occurrence frequency of a word pair (w_i,w_j). A bigram with collocation strength greater than a collocation threshold (computed below) is considered valid and used in the phrase scoring process. Figure 3.10 is an example result from this process.

Word1	Word2	PMI Score
ใน	สาย	7.066089
ความ	รู้	6.066089
เล่ม	นี	12.13218
สูตร	สำเร็จ	7.066089
จัง	เลย	5.481127
เลย	เผือ	5.481127
ต้องการ	เลือน	6.066089
เคล็ดลับ	การ	7.066089
เป็น	ข้าราชการ	15.60615
เลือน	ขัน	6.066089
...	...	...

Figure 3.10 An Example Output Bigram with PMI Score.

$$collocation\ threshold = \log_2(m)$$

where m is the total number of unique words in the entire reviews. After comparing PMI score between Figure 3.10 data and threshold, the result is shown in Figure 3.11.

Word1	Word2	PMI Score
อ่าน	แล้ว	24.36789
หนังสือ	ที	19.9785
ดี	มาก	18.39654
ที	ดี	17.90331
เป็น	หนังสือ	16.71812
เป็น	ข้าราชการ	15.60615
น่า	อ่าน	13.92451
เล่ม	นี	12.13218
อยาก	เป็น	10.81997
แล้ว	อยาก	7.347544
...	...	...

Figure 3.11 Example PMI Scores Greater Than the Threshold

The result is a bigram matrix representation of a directed graph (word graph), as shown in Figure 3.12.

wi/wj	แล้ว	ที่	มาก	ดี	หนังสือ	ข้าราชการ	อ่าน	นี้	เป็น	อยาก
อ่าน	1									
หนังสือ		1								
ดี			1							
ที่				1						
เป็น					1	1				
น่า							1			
เล่ม								1		
อยาก									1	
แล้ว										1

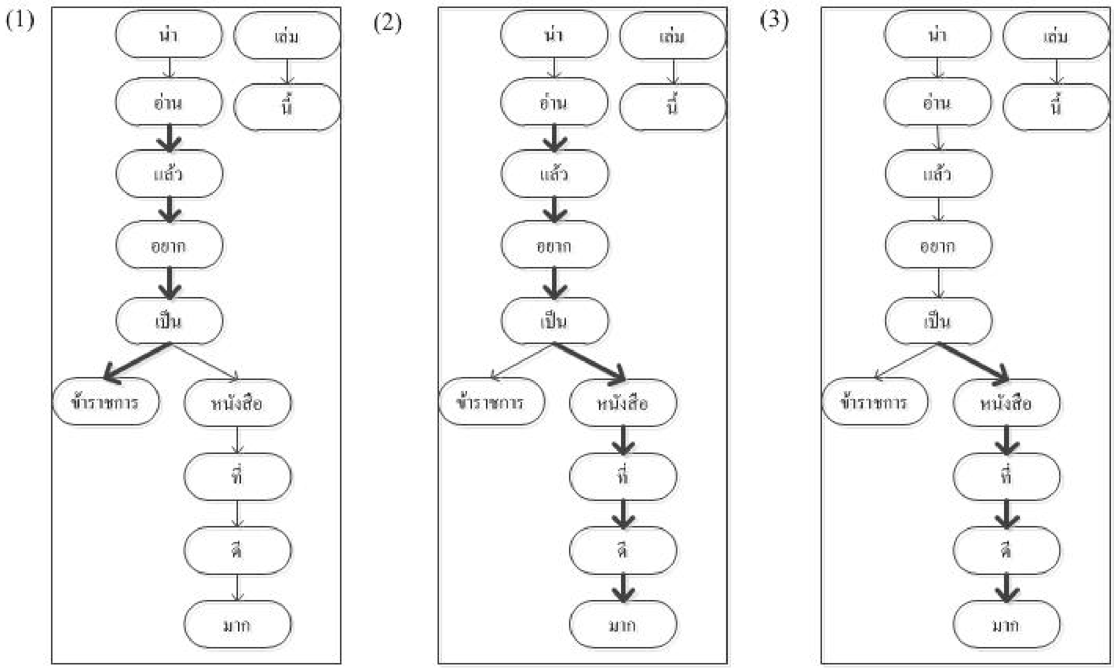

Figure 3.12 Bigram Matrix and the Corresponding Word Graph

In a word graph, a vertex represent a word in the matrix, and a directed edge connecting a word in the row w_i to another word in the column w_j if the bigram (w_i,w_j) has a sufficiently strong collocation strength. A graph is traversed by a modified depth first search (DFS) (Goodrich and Tamassia, 2001: 354-355) where every node takes turn to be the initial vertex, resulting in a set of candidate phrases. The word graph traversal algorithm is shown in following algorithm in Figure 3.13.

Graph traversal was used in Opinosis by grouping the sentences that have very close meaning together, starting with two sentences. The first word in the first sentence is a start node. After creating a link between word1 and word2, the word in the second sentence is added to the graph, and another word in the next sentence is done respectively. After finish creating Opinosis graph, the graph was traversed with depth first search. It searches all parts with selected high score start word with its position less than the average length. During traversing the graph, it will check grammar parallelism and reject incorrect grammar. After that, the graph is selected by highest frequency and linked to candidate phrase, then check valid path, grammar and find summary with score. In Micropinion proposed from group of sentences, the same detail, computation and seed bigram selection applies PMI, map duplicates the seed bigram to create graph. After that, the graph was traversed by depth first search algorithm.

Due to the general ungrammatical nature of the reviews and the inaccuracy of pre-processes in Thai language, our technique is to start traversing a graph at every first word in bigrams; such as อ่าน in (อ่าน, หนัง), as explained in Figure 3.13.

Word Graph Traversal Algorithm
// Input: a bigram graph $G(V,E)$
// Output: a set of candidate phrases S
$S = \{\}$
for $w_i \in$ preceding words in the bigram graph
$S \cup$ {results of a depth first search on G starting at w_i}
end

Figure 3.13 Word Graph Traversal Algorithm

Examples of the phrases generated from the algorithm are shown in Figure 3.14. Each phrase can be seen as a sequence of sub-phrases where a sub-phrase is the longest sequence of words that appear in any review.

เป็นหนังสือที่ดีมาก	อ่านแล้วอยากเป็นหนังสือที่ดีมาก
น่าอ่านแล้วอยากเป็นหนังสือที่ดีมาก	อยากเป็นหนังสือที่ดีมาก
แล้วอยากเป็นหนังสือที่ดีมาก	อ่านแล้วอยากเป็นข้าราชการ
หนังสือที่ดีมาก	น่าอ่านแล้วอยากเป็นข้าราชการ
ที่ดีมาก	แล้วอยากเป็นข้าราชการ
อยากเป็นข้าราชการ	ดีมาก

Figure 3.14 Example of Phrases Resulted from the Word Graph Traversal

Each generated phrase is then scored according to the following equation which takes into account the word importance as signified by the Hopfield network algorithm score, collocation strength between consecutive words, the length of sub-phrases, and is penalized by the mixture of sub-phrases from many reviews.

In addition, we define weight from edge and each word node. The proposed weight for each edge is the strength of two words that were shown together in the document. The collocation is used because it is to combine strong link and frequency of two words together, as well as the importance of the words can be weighed with Hopfield score. Words that have high Hopfield score mean that some other words in the document are relevant with this word; however, we did not use grammar checking because it is too difficult for Thai. Therefore, we propose a method that was generated from our methodology. The phrases that appear to be the same as the original word can be assumed that they are correct phrases and the phrases that had been combined to the long original phrase should be readable. Hence, each phrase that was combined with less original segment has high score and denser phrases should be more important.

$$\text{phrase score} = \frac{\sum_{j=1}^{n}\left[\left\{\sum_{k=1}^{l-1}(collocation(w_k, w_{k+1}) * ((Hopfield(w_k) + Hopfield(w_{k+1})))\right\} * \log_2(l_j)\right]}{l\ * n}$$

where n is the number of sub-phrases combining into the phrase, l_j is the length of sub-phrase j, and l is the length of phrase. The denominator captures actual writing structures in the review texts which promote understandability of the summary without using a grammar or linguistic annotation. An example of the computed phrase score will show in Figure 3.15. Phrases with high scores are selected to generate a summary in the next section.

(1) อ่านแล้วอยากเป็นข้าราชการ

39.8652916861217 = ((43.36+7.34+10.81+15.6)*log(6,2)) /(1*5)

(2) อ่านแล้วอยากเป็นหนังสือที่ดีมาก

18.26803= (((43.36+7.34+10.81)*log(4,2))+
((16.71+19.9+17.9+18.39)*log(5,2)))/(2*8)

(3) เป็นหนังสือที่ดีมาก

42.31714 = ((16.71+19.9+17.9+18.39)*log(5,2))/(1*4)

Figure 3.15 Example of Computed Phase Score

3.3 Generating Abstractive Summary

In this final step, phrases with high scores are grouped according to the following algorithm in Figure 3.16. The similarity between phrases is measured by the cosine similarity, the same as in section 3.1.3. The phrase with the highest score in each group is included in the summary to reduce duplicate information. Effects of the grouping threshold T will be studied in the experiments. The grouping algorithm is described in the following algorithm.

Phrase Grouping Algorithm
//Input: a list of high-scored phrases *S*, ordered by their scores in a descending order
//Output: a set of phrase groups
Assign the first phrase s_1 as the representative for group 1.
for s_i ∈ in *S*
calculate the similarity between s_i and the representative of each existing group s_k.
if (cosine similarity(s_k,s_i) > threshold *T*)
add the item to the corresponding group
recalculate the group representative
else
use s_i to initiate a new group.
end
$s_i = s_{i+1}$
end

Figure 3.16 Phrase-Grouping Algorithm

Figure 3.17 describes clustering algorithm for grouping phrases. The input in this process is to order weight segment and compute the similarity between first high score phrase in the list and all list members.

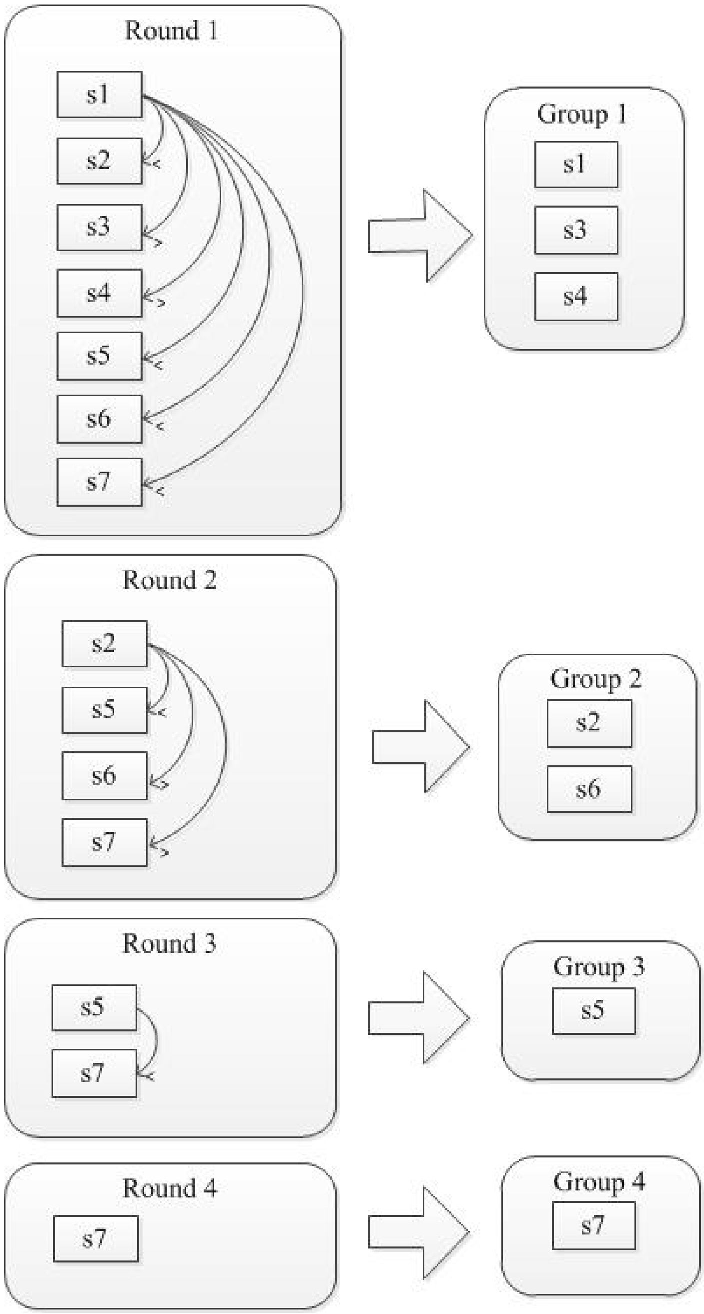

Figure 3.17 Clustering Algorithm for the Summary Candidate Phrases

Therefore, an example of the final output is shown in Figure 3.18.

Global Model	Local Model
เป็นหนังสือที่ดีมาก	เป็นหนังสือที่ดีมาก
อยากเป็นข้าราชการ	อ่านแล้วอยากเป็นข้าราชการ

Figure 3.18 Example of Final Output

CHAPTER 4

RESULTS OF EXPERIMENT

In this section, the proposed technique is evaluated and compared against two baseline methods, using actual customer reviews in Thai.

4.1 Data Set

There is no standard test collection in Thai, especially for opinion summarization. In this research, we gather data from a public online cosmetic website; www.jeban.com, which contains reviews on variety of products from various brands. It is the most popular cosmetic website in Thailand with large number of users. It is also used in a Thai sentiment analysis research (Apisuwankun and Mongkolnavin, 2013). The number of reviews varies upon the popularity of the products. Fifty products were picked randomly, and their reviews are downloaded from the site. For each product, reference summaries were created manually by four Thai female assessors, graduated master’s degree, and are familiar with cosmetic products.

4.2 Evaluation Metrics

In order to evaluate and compare techniques, two types of evaluations are performed: quantitative and qualitative evaluations. Quantitative evaluations intend to measure resemblance between generated summaries and reference (human) summaries. ROUGE (Lin, 2004: 25-26) is popularly used as the main measure for text summarization problems. The measure is based on an n-gram co-occurrence between machine and reference summaries, and is a widely accepted standard for evaluating summarization tasks. In our experiments, we use ROUGE-1, ROUGE-2 and ROUGE-

SU4 measure. ROUGE-1 and ROUGE-2 have shown to have most correlation with human summaries (Filippova, 2010: 322-330), while higher order ROUGE-N scores (N > 1) estimates the fluency of summaries. ROUGE can be calculated as follows:

$$ROUGE_n(X) = \frac{\sum_{j=1}^{h} \sum_{i \in N_n} min(X_n(i), M_n(i,j))}{\sum_{j=1}^{h} \sum_{i \in N_n} M_n(i,j)}$$

where N_n represents the set of all n-grams, and i is one member from N_n. $X_n(i)$ is the number of times the n-gram i occurred in the summary, and $M_n(i,j)$ is the number of times the n-gram i occurred in the j-the reference summary. There are totally h reference summaries. When computing ROUGE score for a summarization system, reference summaries are created in advance. In our experiments, for each product, there were four reference summaries. All reference summaries were combined together where only one phrase was kept for a redundant meaning, i.e., h = 1. ROUGE-1, ROUGE-2, and ROUGE-SU4, the DUC automatic evaluation criteria (Eduard, Lin, Zhou and Fukumoto, 2006:1-4), are used to compare our systems built upon the proposed concept. ROUGE-2 evaluates a system summary by matching its bigrams with the reference summaries. ROUGE-SU4 matches unigrams while skip-bigrams of the summary with reference summaries, where the skipped bigram is a pair of words in their sentence order, allowing gaps within a limited size.

For qualitative evaluations, three dimensions are measured: informative, grammatical, and non-redundancy aspects of the summary. The informative aspect measures how much users can learn from the summary, the grammatical aspect measures readability of the summary and the non-redundancy aspect measures the uniqueness of phrases without unnecessary repetitions of facts in the summary. Each of these aspects is given a score from 1 (minimum) to 3 (maximum) by human assessors. Two baseline models are used to compare with the proposed technique. Both models generate extractive summaries.

Baseline 1: The text segments selected by the segment extraction process are grouped, and the phrase with the highest phrase score from each group is included in the summary.

Baseline 2: The text segments selected by the segment extraction process are examined by human assessors. Only segments with non-redundant meaning are included in the summary.

Baseline 1 is a classic extractive text summarization method, while in Baseline 2 humans are more involved in the process with helps from the machine.

4.3 Results

The majority of reviews downloaded from the website use informal language and are short phrases; however, they are understandable by readers in their contexts; for example, ไม่มัน (not oily), หน้าไม่มัน (not oily face), ไม่คุมมัน (not oil control), etc. In the proposed technique, two models are studied: local and global models.

Local model: the term collocation statistics are calculated from the extracted segments.

Global model: The term collocation statistics are calculated from the entire set of customer reviews.

In this Chapter, first we study the parameters that affect the performance of the proposed techniques. Then, we quantitatively evaluate the models and compare them with the 2 baselines. Finally, we evaluate all the models in qualitative aspects.

4.3.1 Effects of Model Parameters

Two important parameters that can affect the performance of the proposed technique are the top R percent of segments to be extracted from the reviews and the grouping threshold, T. We randomly sample 30% of the 50 products used in our study and vary the combinations of the two parameters, i.e., $R \in \{10\%, 20\%, \ldots, 90\%\}$ and $T \in \{0.1, 0.2, \ldots, 0.9\}$, a nd calculate the average F-measures over the sample. The results show that the combination of R = 20% and T = 0.5 yields the highest value of F-measure in all three ROUGE measures. Table 4.1 shows the performance of the local model with T = 0.5 and R varied from 10% to 90%. The global model was not affected by *R*. We can see that the model performs best when the top 20% of the segments are extracted, while at other values the model does not perform equally well.

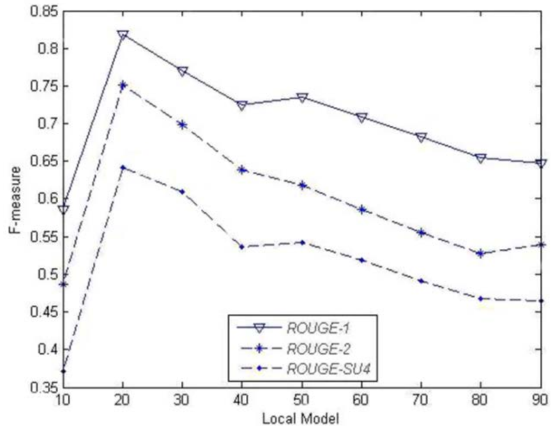

Figure 4.1 Effects of Segment Subset Sizes on the Local Model (T is set at 0.5)

Figure 4.2 shows the effects of varying the values of the grouping threshold T from 0.1 to 0.9, when R is set at 20%. We can see that in all three ROUGE measures, the threshold of 0.5 yields the highest performance compared to other values in general. Thus, in further experiments, we will set the values of parameters R and T at 20% and 0.5, respectively.

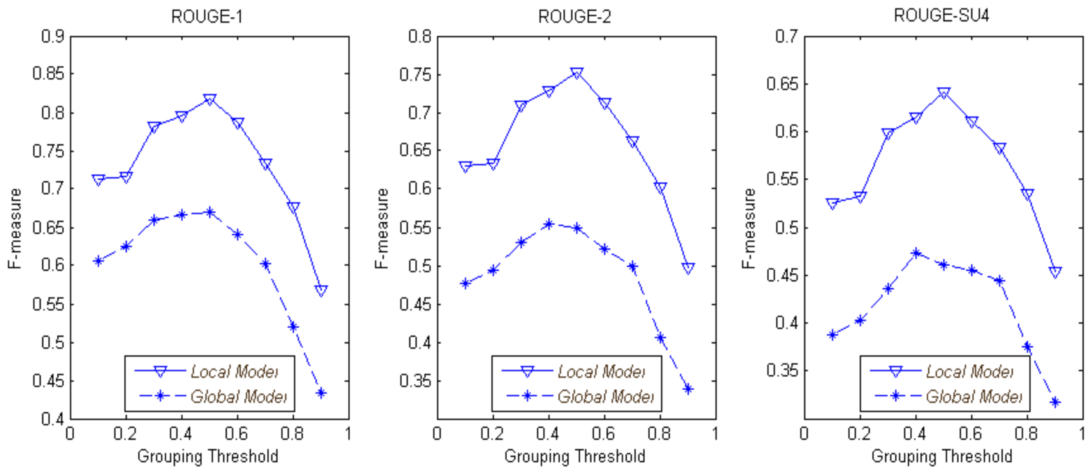

Figure 4.2 Effects of the Grouping Threshold T (R is set at 20%)

4.3.2 Quantitative Evaluations

We studied the performance of the proposed models (the local and global models) and compared them with the two baseline models quantitatively, using the three standard ROGUE measures. The results are shown in Table 4.1.

Table 4.1 Performance of the Proposed Models and the Baseline Models

	Recall			Precision			F-measure		
	Rouge-1	Rouge-2	Rouge-SU4	Rouge-1	Rouge-2	Rouge-SU4	Rouge-1	Rouge-2	Rouge-SU4
Local Model	0.6882	0.5620	0.4480	0.9014	0.9238	0.9474	0.7583	0.6708	0.5737
Global Model	0.5804	0.4262	0.3387	0.7945	0.8291	0.8661	0.6511	0.5395	0.4619
Baseline1	0.7321	0.4667	0.4557	0.5604	0.8099	0.7644	0.6089	0.5559	0.5332
Baseline2	0.7058	0.4884	0.4676	0.5132	0.6506	0.6246	0.5686	0.5181	0.4870

According to the table, we can see that the local model is more effective than the global model in every measure. This shows the effectiveness of the segment extraction process where a representative set of segments is selected from the reviews. In terms of recall, the baseline models have higher recall than the models in ROUGE-1, since baseline models returns the actual segments from the review texts which contain large number of single words related to the proposed model; however, it will affect the precision as shown in the table. In other ROUGE measures, which are based on bigrams and skip-bigrams, the local model yields high recall. In terms of precision, both proposed models are more effective than other baselines, as discussed above. In addition, Baseline 1 performs better than Baseline 2 in almost all measures, showing the effectiveness of the phrase scoring criteria. Overall, the local model performs the best quantitatively.

4.3.3 Qualitative Evaluation

The quantitative evaluations in the previous section did not show the quality of summary in the eyes of the readers. The three qualitative dimensions studied in this research include informative, grammatical, and non-redundancy aspects of the summary, whose score is provided by human assessors. The results are shown in Table 4.2. In the informative aspect, both proposed models provide more information to readers while the baseline models lack some points from the reviews. In the grammatical aspect, the local model is shown to generate more readable summary than the global model. Since review texts are understandable by human, and the baseline models take actual phrases from the reviews, grammatical aspects are not applicable to those models. In the non-redundancy aspect, Baseline 1 performs well because the phrases are selected from the phrase scoring formula, which contains less redundancy than Baseline 2. However, the local model is found to perform best in this non-redundancy aspect. We can see that the local model produces highest quality summary among the models studied.

Table 4.2 Quality of the Summaries Generated by Different Models

	Baseline1	Baseline2	Local Model	Global Model
Informative	1.8235	1.7235	2.2353	2.1176
Grammatically	N/A	N/A	2.1176	2.0588
Non-redundancy	2.1176	1.8592	2.2235	2.1471

In addition, Table 4.3 shows the details of human assessors' score allocations in all aspects of the qualitative evaluations. For the local model, the majority (41.18%) of assessors gave moderate amount of information, while 35.29% of them gave the highest amount of information to readers. For the global model, large portion of the assessor sees that the summaries generated are not informative. In grammatical aspect, with the local model, the proportion of people giving the moderate degree are the same as that giving the highest degree, while with the global model, most people think that its grammatical performance is moderate. In the non-redundancy aspect, every model produces reasonably unique summaries, and more inclined to the high quality side than to the low quality one.

From all the evaluations, our proposed local model is found to produce the best summary across different measures in both quantitative and qualitative aspects.

Table 4.3 Quality Score Allocations in Different Aspects by Human Assessors

System	Informative			Grammatical			Non-Redundancy		
	1	2	3	1	2	3	1	2	3
Baseline1	23.53%	41.18%	35.29%	N/A	N/A	N/A	20.34%	58.24%	21.32%
Baseline2	28.58%	40.10%	31.32%	N/A	N/A	N/A	17.65%	58.82%	23.53%
Local Model	23.53%	35.29%	41.18%	17.65%	41.18%	41.18%	17.65%	52.94%	29.41%
Global Model	47.06%	41.18%	11.76%	23.53%	41.18%	35.29%	29.41%	35.29%	35.29%

CHAPTER 5

CONCLUSION

Automatically summarizing a large number of customer reviews is challenging. The task is more complicated for reviews written in Thai. Thai words are written continuously without space, there is no s ymbol to signify the end of a sentence, and many reviews are written informally and grammatically incorrect. This makes it infeasible for a method to rely on a ccurate word identification and its linguistic function. Two general types of text summarization consist of extractive summarization where the summary is a s et of actual sentences or phrases from the reviews, and abstractive summarization where the summary is rewritten from the contents of the reviews. This approach is more difficult and therefore less popular than the extractive approach.

In this research, an automatic, abstractive opinion summarization technique for reviews written in Thai is presented. The technique begins with an extraction of important text segments representing opinions of customers using a graph-ranking algorithm. Bigrams are generated from the segments, and strong relationships among them are used to create the graph. Through a traversal, a set of text phrases are created and ranked according to word importance, collocations among words, and structures of review texts. Highly similar phrases are grouped, and the top-ranked phrase in each group is included in the summary. The technique is fully unsupervised, domain independent, while grammar and linguistic annotation are excluded and relying only on important text segments extracted, redundancies and writing structures in the review texts. Two models, which are part of the proposed technique, are studied: the local and the global models. Their major difference is in the bigram matrix construction, which leads to different phrases summary for generating. Both models are evaluated and compared with 2 extractive baseline methods quantitatively, using 3 standard text summarization measures, i.e., ROUGE-1, ROUGE-2, and ROUGE-SU4, and qualitatively using three aspects that are informative, grammatical, and non-redundancy. The results show that the local model generates the summaries that resemble human summaries the most, as measured by ROUGE and has the highest quality when measured by the three qualitative measures.

In future work, the Hopfield network scores of segments can be included in the calculation of the phrase score to give higher weights for terms from important segments.

The proposed phrase scoring formula and its weightings were from an analysis of and tailored for Thai opinions. It may not directly by applicable to other languages. To use with another language, the phrase scoring formula may need to be modified to suit the characteristics of the language. Other parts of the proposed technique can be applied to opinions in other languages as proposed.

BIBLIOGRAPHY

Apisuwankun, P. and Mongkolnavin, J. 2013. **Opinion Strength Identification in Customer Review Summarizing System Using Association Rule Technique.** Presented at The International Conference on E-Technologies and Business on the Web (EBW 2013). Pp. 16-21.

Baeza-Yates, R.; Araújo Neto Ribeiro, B. and Ribeiro-Neto, B. 1999. **Modern Information Retrieval.** Boston, MA.: Addison-Wesley.

Banko, M.; Mittal, V.; Kantrowitz, M. and Goldstein, J. 1999. Generating Extraction-Based Summaries from Hand-Written Summaries by Aligning Text Spans. In **Proceedings of the Pacific Association for Computational Linguistics PACLING.** Pp. 1-5.

Bheganan, P.; Nayak, R. and Xu, Y. 2009. Thai Word Segmentation with Hidden Markov Model and Decision Tree. In **Advances in Knowledge Discovery and Data Mining. T. Theeramunkong, B. Kijsirikul, N. Cercone and T.-B. Ho eds.** Springer Berlin Heidelberg. Pp. 74-85.

Brin, S. and Page, L. 1998. The Anatomy of a Large-scale Hypertextual Web Search Engine. In **Proceedings of the Seventh International Conference on World Wide Web 7.** Amsterdam: Elsevier Science Publishers. Pp. 107-117.

Carenini, G. and Cheung, J. C. K. 2008. Extractive vs. NLG-based Abstractive Summarization of Evaluative Text: The Effect of Corpus Controversiality. In **Proceedings of the Fifth International Natural Language Generation Conference.** Stroudsburg, PA: Association for Computational Linguistics. Pp. 33-41.

Charoenpornsawat, P. and Sornlertlamvanich, V. 2001. Automatic Sentence Break Disambiguation for Thai. In **Proceedings of International Conference on Computer Processing of Oriental Languages (ICCPOL).** Pp. 231-235.

Chen, H. and Ng, T. 1995. An Algorithmic Approach to Concept Exploration in a Large Knowledge Network (Automatic Thesaurus Consultation): Symbolic Branch-and-Bound Search vs. Connectionist Hopfield Net Activation. **Journal of the American Society for Information Science.** 46 (June): 348-369.

Damani, O. P. 2013. Improving Pointwise Mutual Information (PMI) by Incorporating Significant Co-occurrence. In **Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing.** Seattle, Washington. Pp. 163-169.

Eduard, H.; Lin, C.; Zhou, L. and Fukumoto, J. 2006. Automated Summarization Evaluation with Basic Elements. In **Proceedings of the Fifth Conference on Language Resources and Evaluation (LREC).** Genoa, Italy. Pp. 1-4.

Filippova, K. 2010. Multi-sentence Compression: Finding Shortest Paths in Word Graphs. In **Proceedings of the 23rd International Conference on Computational Linguistics.** Stroudsburg, PA: Association for Computational Linguistics. Pp. 322-330.

Ganesan, K.; Zhai, C. and Han, J. 2010. Opinosis: A Graph-based Approach to Abstractive Summarization of Highly Redundant Opinions. In **Proceedings of the 23rd International Conference on Computational Linguistics.** Stroudsburg, PA: Association for Computational Linguistics. Pp. 340-348.

Ganesan, K.; Zhai, C. and Viegas, E. 2012. Micropinion Generation: An Unsupervised Approach to Generating Ultra-concise Summaries of Opinions. In **Proceedings of the 21st International Conference on World Wide Web.** New York, NY: ACM. Pp. 869-878.

Goodrich, M. T. and Tamassia, R. 2001. **Algorithm Design: Foundations, Analysis, and Internet Examples.** New York: John Wiley & Sons.

Ishikawa, K.; ANDO, S.; Doi, S. and Okumura, A. 2002. **Trainable Automatic Text Summarization Using Segmentation of Sentence.** Retrieved April 16, 2014. from research.nii.ac.jp/ntcir/NTCIR3-TSC-IshikawaK.pdf

Jeban.com. 2014. **Revlon - Colorstay™ Makeup**. Retrieved April 16, 2014. from http://www.jeban.com/reviews_product.php?p_id=18768&page=1.

Kleinberg, J. M. 1999. Authoritative Sources in a Hyperlinked Environment. **Journal of the ACM (JACM).** 46 (September): 604-632.

Lin, C., 2004. Rouge: A Package for Automatic Evaluation of Summaries. In **Proceedings of the Workshop on Text Summarization Branches Out (WAS 2004).** Barcelona, Spain, Pp. 25-26.

Liu, R.; An, Y. and Song, L. 2011. Chinese Multi-document Summarization Based on Opinion Similarity. **Procedia Engineering.** 15 (2011): 2026-2031.

Lloret, E. and Palomar, M. 2011. **Analyzing the Use of Word Graphs for Abstractive Text Summarization.** The First International Conference on Advances in Information Mining and Management. Barcelona, Spain. Pp. 61-66.

Lloret, E.; Romá-Ferri, M. T. and Palomar, M. 2013. COMPENDIUM: A Text Summarization System for Generating Abstracts of Research Papers. **Data & Knowledge Engineering.** 88 (November): 164-175.

Luhn, H. P. 1958. The Automatic Creation of Literature Abstracts. **IBM Journal of Research and Development.** 2 (April): 159-165.

Makeupalley.com. 2014. **Liquid – Revlon-ColorStay Makeup Discontinued and Reformulated Early 2013.** Retrieved April 16, 2014. from http://www.makeupalley .com/product/showreview.asp/ ItemId=13842/ ColorStay-Makeup- Discontinued-and-Reformulated-Early-2013]/ Revlon/Liquid.

Salton, G.; Allan, J.; Buckley, C. and Singhal, A. 1999. Automatic Analysis Theme Genertion and Summarization of Machine-readable Text. In **Readings in Information Visualization.** K. Card, ed. San Francisco, Calif: Morgan Kaufmann Publishers. Pp. 413-418.

Shen, D.; Chen, Z.; Yang, Q.; Zeng, H.-J.; Zhang, B.; Lu, Y. and Ma, W.-Y. 2004. Web-page Classification Through Summarization. In **Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.** New York, NY: ACM. Pp. 242-249.

Sornil, O. and Gree-ut, K. 2006. An Automatic Text Summarization Approach using Content-Based and Graph-Based Characteristics. In **IEEE Conference on Cybernetics and Intelligent Systems.** Bangkok, Thailand: IEEE. Pp. 1-6.

BIOGRAPHY

NAME	Orawan Chaowalit
ACADEMIC BACKGROUND	Bachelor of Science in Computer science from Silpakorn University, Nakorn Pathom, Thailand in 1997. Master of Science in Computer Science from Thammasat University, Bangkok, Thailand in 2004
PRESENT POSITION	Lecturer in Department of Computing, Department of Science from Silpakorn Univeristy, Nakorn Pathom, Thailan
EXPERIENCES	1997 - 2004 Programmer at Computer Centre, Silpakorn Universitiy, Nakorn Pathom,Thailand Received a scholarship from Office of The Higher Education Commission for study master and doctoral degree.
PUBLICATION	Orawan Chaowalit and Ohm Sornil, 2014. An Automatic Approach to Generating Abstractive Summary for Thai Opinions. **In International Journal of Advancements in Computing Technology (IJACT)**.