A Just-In-Time Keyword Extraction from Meeting Transcripts

Hyun-Je Song **Junho Go** **Seong-Bae Park** **Se-Young Park**
School of Computer Science and Engineering
Kyungpook National University
Daegu, Korea
{hjsong,jhgo,sbpark,sypark}@sejong.knu.ac.kr

Abstract

In a meeting, it is often desirable to extract keywords from each utterance as soon as it is spoken. Thus, this paper proposes a just-in-time keyword extraction from meeting transcripts. The proposed method considers two major factors that make it different from keyword extraction from normal texts. The first factor is the temporal history of preceding utterances that grants higher importance to recent utterances than old ones, and the second is topic relevance that forces only the preceding utterances relevant to the current utterance to be considered in keyword extraction. Our experiments on two data sets in English and Korean show that the consideration of the factors results in performance improvement in keyword extraction from meeting transcripts.

1 Introduction

A meeting is generally accomplished by a number of participants and a wide range of subjects are discussed. Therefore, it would be helpful to meeting participants to provide them with some additional information related to the current subject. For instance, assume that a participant is discussing a specific topic with other participants at a meeting. The summary of previous meetings on the topic is then one of the most important resources for her discussion.

In order to provide information on a topic to participants, keywords should be first generated for the topic since keywords are often representatives of a topic. A number of techniques have been proposed for automatic keyword extraction (Frank et al., 1999; Turney, 2000; Mihalcea and Tarau, 2004; Wan et al., 2007), and they are designed to extract keywords from a written document. However, they are not suitable for meeting transcripts. In a meeting, it is often desirable to extract keywords at the time at which a new utterance is made for just-in-time service of additional information. Otherwise, the extracted keywords become just the important words at the end of the meeting.

Two key factors for just-in-time keyword extraction from meeting transcripts are time of preceding utterances and topic of current utterance. First, current utterance is affected by temporal history of preceding utterances. That is, when a new utterance is made it is likely to be related more closely with latest utterances than old ones. Second, the preceding utterances which carry similar topics to current utterance are more important than irrelevant utterances. Since a meeting consists of several topics, the utterances that have nothing to do with current utterance are inappropriate as a history of the current utterance.

This paper proposes a graph-based keyword extraction to reflect these factors. The proposed method represents an utterance as a graph of which nodes are candidate keywords. The preceding utterances are also expressed as a history graph in which the weight of an edge is the temporal importance of the keywords connected by the edge. To reflect the temporal history of utterances, *forgetting curve* (Wozniak, 1999) is adopted in updating the weights of edges in the history graph. It expresses effectively not only the reciprocal relation between memory re-

Proceedings of NAACL-HLT 2013, pages 888–896,
Atlanta, Georgia, 9–14 June 2013.

tention and time, but also active recall that makes frequent words more consequential in keyword extraction. Then, a subgraph that is relevant to the current utterance is derived from the history graph, and used as an actual history of the current utterance. The keywords of the current utterance are extracted by TextRank (Mihalcea and Tarau, 2004) from the merged graph of the current utterance and the history graphs.

The proposed method is evaluated with two kinds of data sets: the National Assembly transcripts in Korean and the ICSI meeting corpus (Janin et al., 2003) in English. The experimental results show that it outperforms both the TFIDF framework (Frank et al., 1999; Liu et al., 2009) and the PageRank-based graph model (Wan et al., 2007). One thing to note is that the proposed method improves even the supervised methods that do not reflect utterance time and topic relevance for the ICSI corpus. This proves that it is critical to consider time and content of utterances simultaneously in keyword extraction from meeting transcripts.

The rest of the paper is organized as follows. Section 2 reviews the related studies on keyword extraction. Section 3 explains the overall process of the proposed method, and Section 4 addresses its detailed description how to reflect meeting characteristics. Experimental results are presented in Section 5. Finally, Section 6 draws some conclusions.

2 Related Work

Keyword extraction has been of interest for a long time in various fields such as information retrieval, document clustering, summarization, and so on. Thus, there have been many studies on automatic keyword extraction. The frequency-based keyword extraction with TFIDF weighting (Frank et al., 1999) and the graph-based keyword extraction (Mihalcea and Tarau, 2004) are two base models for this task. Many studies recently tried to extend them by incorporating specific information such as linguistic knowledge (Hulth, 2003), web-based resource (Turney, 2003), and semantic knowledge (Chen et al., 2010). As a result, they show good performance on written text. However, it is difficult to use them directly for spoken genres, since spoken genres have significantly different characteristics from written text.

There have been a few studies focused on keyword extraction from spoken genres. Among them, the extraction from meetings has attracted more concern, since the need for grasping important points of a meeting or an opinion of each participant has increased. The studies on meetings focused on the exterior features of meeting dialogues such as unstructured and ill-formed sentences. Liu et al. (2009) used some knowledge sources such as Part-of-Speech (POS) filtering, word clustering, and sentence salience to reflect dialogue features, and they found out that a simple TFIDF-based keyword extraction using these knowledge sources works reasonably well. They also extended their work by adopting various features such as decision making sentence features, speech-related features, and summary features that reflect meeting transcripts better (Liu et al., 2011). Chen et al. (2010) extracted keywords from spoken course lectures. In this study, they considered prosodic information from HKT forced alignment and topics in a lecture generated by Probabilistic Latent Semantic Analysis (pLSA). These studies focused on the exterior characteristics of spoken genres, since they assumed that entire scripts are given in advance and then they extracted keywords that best describe the scripts. However, to the best of our knowledge, there is no previous study considered time of utterances which is an intrinsic element of spoken genres.

The relevance between current utterance and preceding utterances is also a critical feature in keyword extraction from meeting transcripts. The study that considers this relevance explicitly is *CollabRank* proposed by Wan and Xiao (2008). This is collaborative approach to extract keywords in a document. In this study, it is assumed that a few neighbor documents close to a current document can help extract keywords. Therefore, they applied a clustering algorithm to a document set and then extracted words that are reinforced by the documents within a cluster. However, this method also does not consider the utterance time, since it is designed to extract keywords from normal documents. As a result, if it is applied to meeting transcripts, all preceding utterances would affect the current utterance uniformly, which leads to a poor performance.

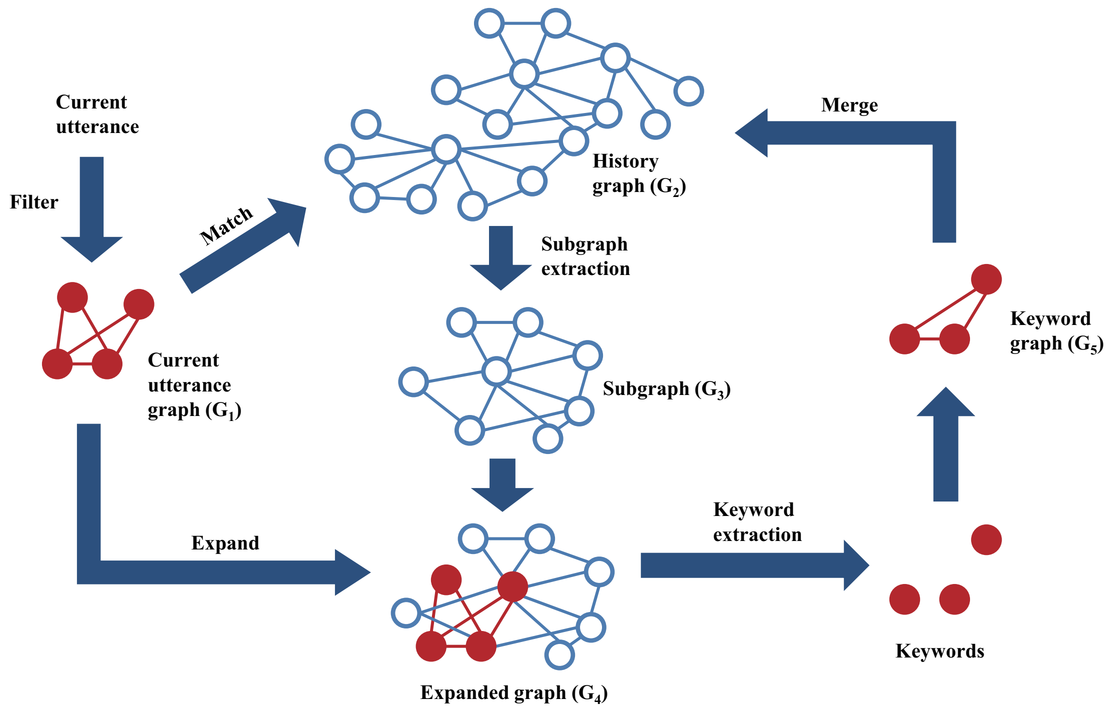

Figure 1: The overall process of the just-in-time keyword extraction from meeting transcripts.

3 Just-In-Time Keyword Extraction for a Meeting

Figure 1 depicts the overall process of extracting keywords from an utterance as soon as it is spoken. We represent all the components in a meeting as graphs. This is because graphs are effective to express the relationship between words, and the graph operations that are required for keyword extraction are also efficiently performed. That is, whenever an utterance is spoken, it is represented as a graph (G_1) of which nodes are the potential keywords in the utterance. This graph is named as *current utterance graph*.

The summary of all preceding utterances is also represented as a *history graph* (G_2). We assume that only the preceding utterances that are directly related with the current utterance are important for extracting keywords from the current utterance. Therefore, a subgraph of G_2 that maximally covers the current utterance graph (G_1) is extracted. This subgraph is labeled as G_3 in Figure 1. Then, the current utterance graph G_1 is expanded by merging it and G_3. This expanded graph (G_4) is a combined representation of the current and preceding utterances, and then the keywords of the current utterance is extracted from this graph. The keywords are so-called hub nodes of G_4.

After keywords are extracted from the current utterance, the current utterance becomes a part of the history graph for the next utterance. For this, the extracted keywords are also represented as a graph (G_5), and it is merged into the current history G_2. This merged graph becomes a new history graph for the next utterance. In merging two graphs, the weight of each edge in G_2 is updated to reflect the temporal history. If an edge is connecting two nouns from an old utterance, its weight becomes small. In the same way, the weights for the edges from recent utterances get large. The weights of the edges from G_5 are 1, the largest possible value.

4 Graph Representation and Weight Update

4.1 Current Utterance Graph and History Graph

Current utterance graph is a graph-representation of the current utterance. When current utterance consists of m words, we first extract the potential keywords from the current utterance. Since all words within the current utterance are not keywords, some words are filtered out. For this filtering out, we follow the POS filtering approach proposed by Liu et al. (2009). This approach filters out non-keywords using a stop-word list and POS tags of the words. Assume that n words remain after the filtering out, where $n \leq m$. These n words become the vertices of the current utterance graph.

Formally, the *current utterance graph* $G_1 = (V_1, E_1)$ is an undirected graph, where $|V_1| = n$. E_1 is a set of edges and each edge implies that the nouns connected by the edge co-occur within a window sized W. For each $e^1_{ij} \in E_1$ that connects nodes v^1_i and v^1_j, its weight is given by

$$w^1_{ij} = \begin{cases} 1 & \text{if } v^1_i \& v^1_j \text{ cooccur within the window,} \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

In a meeting, preceding utterances affect the current utterance. We assume that only the keywords of preceding utterances are effective. Therefore, the *history graph* $G_2 = (V_2, E_2)$ is an undirected graph of keywords in the preceding utterances. That is, all vertices in V_2 are keywords extracted from one or more previous utterances, and the edge between two keywords implies that they co-occurred at least once. Every edge in E_2 has a weight that represents its temporal importance.

The history graph is updated whenever keywords are extracted from a new utterance. This is because the current utterance becomes a part of the history graph for the next utterance. As a history, old utterances are less important than recent ones. Thus, the temporal importance should decrease gradually according to the passage of time. In addition, the keywords which occur frequently at a meeting are more important than those mentioned just once or twice. Since the frequently-mentioned keywords are normally major topics of the meeting, their influence should last for a long time.

To model these characteristics, the *forgetting curve* (Wozniak, 1999) is adopted in updating the history graph. It models the decline of memory retention in time. Figure 2 shows a typical representation of the forgetting curve. The X-axis of this figure is time and the Y-axis is memory retention. As shown in this figure, memory retention of new information decreases gradually by the exponential nature of forgetting. However, whenever the information is repeated, it is recalled longer. This is formulated as

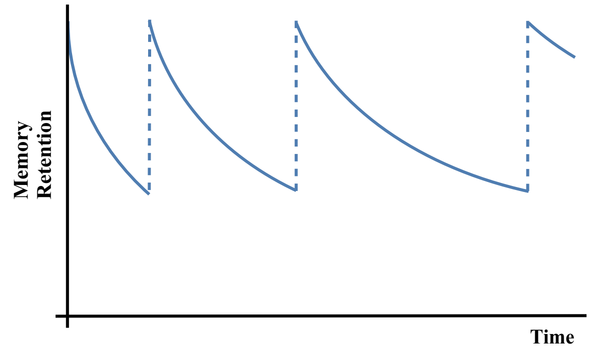

Figure 2: Memory retention according to time.

$$R = e^{-\frac{t}{S}},$$

where R is memory retention, t is time, and S is the relative strength of memory.

Based on the forgetting curve, the weight of each edge $e^2_{ij} \in E_2$ between keywords v^2_i and v^2_j is set as

$$w^2_{ij} = \exp^{-\frac{t}{f(v_i, v_j)}}, \quad (2)$$

where t is the elapse of utterance time and $f(v_i, v_j)$ is the frequency that v_i and v_j co-occur from the beginning of the meeting to now. According to this equation, the temporal importance between keywords decreases gradually as time passes by, but the keyword relations repeated during the meeting are remembered for a long time in the history graph.

4.2 Keyword Extraction by Merging Current Utterance and History Graphs

All words within the history graph are not equally important in extracting keywords from the current utterance. In general, many participants discuss a wide range of topics in a meeting. Therefore, some preceding utterances that shares topics with the current utterance are more significant. We assume that the preceding utterances that contain the nouns in the current utterance share topics with the current utterance. Thus, only a subgraph of G_2 that contain words in G_1 is relevant for keyword extraction from G_1.

Given the current utterance graph $G_1 = (V_1, E_1)$ and the history graph $G_2 = (V_2, E_2)$, the relevant graph $G_3 = (V_3, E_3)$ is a subgraph of G_2. Here, $V_3 = (V_1 \cap V_2) \cup adjacency(V_1)$ and $adjacency(V_1)$ is a set of vertices from G_2 which are directly connected to the words in V_1. That is, V_3 contains the words of G_1 and their direct neighbor words in G_2. E_3 is a subset of E_2. Only the edges that appear in E_2 are included in E_3. The weight w_{ij}^3 of each $e_{ij}^3 \in E_3$ is also borrowed from G_2. That is, $w_{ij}^3 = w_{ij}^2$. Therefore, G_3 is a 1-walk subgraph[1] of G_2 in which words in G_1 and their neighbor words appear.

The keywords of the current utterance should reflect the relevant history as well as the current utterance itself. For this purpose, G_1 is expanded with respect to G_3. The expanded graph $G_4 = (V_4, E_4)$ of G_1 is defined as

$$V_4 = V_1 \cup V_3,$$
$$E_4 = E_1 \cup E_3.$$

For each edge $e_{ij}^4 \in E_4$, its weight w_{ij}^4 is determined to be the larger value between w_{ij}^1 and w_{ij}^3 if it appears in both G_1 and G_3. When it appears in only one of the graphs, w_{ij}^4 is set to be the weight of its corresponding graph. That is,

$$w_{ij}^4 = \begin{cases} \max(w_{ij}^1, w_{ij}^3) & \text{if } e_{ij}^4 \in E_1 \text{ and } e_{ij}^4 \in E_3, \\ w_{ij}^1 & \text{if } e_{ij}^4 \in E_1 \text{ and } e_{ij}^4 \notin E_3, \\ w_{ij}^3 & \text{otherwise.} \end{cases}$$

From this expanded graph G_4, the keywords are extracted by TextRank (Mihalcea and Tarau, 2004). TextRank is an unsupervised graph-based method for keyword extraction. It singles out the key vertices of a graph by providing a ranking mechanism. In order to rank the vertices, it computes the score of each vertex $v_i^4 \in V_4$ by

$$S(v_i^4) = (1-d) + d \cdot \sum_{v_j^4 \in adj(v_i^4)} \frac{w_{ji}^4}{\sum_{v_k^4 \in adj(v_j^4)} w_{jk}^4} S(v_j^4), \quad (3)$$

where $0 \leq d \leq 1$ is a damping factor and $adj(v_i)$ denotes v_i's neighbors. Finally, the words whose score is larger than a specific threshold θ are chosen as keywords. Especially when the current utterance is the first utterance of a meeting, the history graph does not exist. In this case, the current utterance graph becomes the expanded graph ($G_4 = G_1$), and keywords are extracted from the current utterance graph.

The proposed method extracts keywords whenever an utterance is spoken. Thus, it tries to extract keywords even if the current utterance is not related to the topics of a meeting or is too short. However, if the current utterance is irrelevant to the meeting, it has just a few connections with other previous utterances, and thus the potential keywords in this utterance are apt to have a low score. The proposed method, however, does not select the words whose score is smaller than the threshold θ as keywords. As a result, it extracts only the relevant keywords during the meeting.

Since the keywords for the current utterance should be the history for the next utterance, they have to be reflected into the history graph. Therefore, a *keyword graph* $G_5 = (V_5, E_5)$ is constructed from the keywords. Here, V_5 is a set of keywords extracted from G_4, and E_5 is a subset of E_4 that corresponds to V_5. The weights of edges in E_5 are same with those in E_4. That is, $w_{ij}^5 = w_{ij}^4$. The keyword graph G_5 is then merged into the history graph G_2 in the same way that G_1 and G_3 are merged. As stated above, the weights of the edges in the history graph G_2 are updated by Equation (2). Therefore, before merging G_5 and G_2, all weights of G_2 are updated by increasing t as $t+1$ to reflect temporal importance of preceding utterances.

5 Experiments

The proposed method is evaluated with two kinds of data sets: the *National Assembly transcripts* in Korean and the *ICSI meeting corpus* in English. Both data sets are the records of meetings that are manually dictated by human transcribers.

[1] If a m-walk subgraph ($m > 1$) is used, more affluent history is used. However, this graph contains some words irrelevant to the current utterance. According to our experiments, 1-walk subgraph outperforms other m-walk subgraphs where $m > 1$. In addition, extracting G_3 becomes expensive for large m.

Table 1: Simple statistics of the National Assembly transcripts

	the first meeting	the second meeting
No. of utterances	1,280	573
Average No. of words per utterance	7.22	10.17

5.1 National Assembly Transcripts in Korean

The first corpus used to evaluate our method is the National Assembly transcripts[2]. This corpus is obtained from the Knowledge Management System of the National Assembly of KoreaIt is transcribed from the 305th assembly record of the Knowledge Economy Committee in 2012. Table 1 summarizes simple statistics of the National Assembly transcripts. The 305th assembly record actually consists of two meetings. The first meeting contains 1,280 utterances and the second has 573 utterances. The average number of words per utterance in the first meeting is 7.22 while the second meeting contains 10.17 words per utterance on average. The second meeting transcript is used as a development data set to determine window size W of Equation (1), the damping factor d of Equation (3), and the threshold θ. For all experiments below, d is set 0.85, W is 10, and θ is 0.28. The remaining first meeting transcript is used as a data set to extract keywords since this transcript contains more utterances. Only nouns are considered as potential keywords. That is, only the words whose POS tag is NNG (common noun) or NNP (proper noun) can be a keyword.

Three annotators are engaged to extract keywords manually for each utterance in the first meeting transcript, since the Knowledge Management System does not provide the keywords[3]. The average number of keywords per utterance is 2.58. To see the inter-judge agreement among the annotators, the Kappa coefficient (Carletta, 1996) was investigated. The kappa agreement of the National Assembly transcript is 0.31 that falls under the category of 'Fair'. Even though all congressmen in the transcript belong to the same committee, they discussed various topics at the meeting. As a result, the keywords are difficult to be agreed unanimously by all three annotators. Therefore, in this paper the words that are recommended by more than two annotators are chosen as keywords.

The evaluation is done with two metics: F-measure and the weighted relative score (WRS). Since the previous work by Liu et al. (2009) reported only F-measure and WRS, F-measure instead of precision/recall are used for the comparison with their method. The weighted relative score is derived from *Pyramid* metric (Nenkova and Passonneau, 2004). When a keyword extraction system generates keywords which many annotators agree, a higher score is given to it. On the other hand, a lower score is given if fewer annotators agree.

The proposed method is compared with two baseline models to see its relative performance. One is the frequency-based keyword extraction with TFIDF weighting (Frank et al., 1999) and the other is TextRank in which the weight of edges is mutual information between vertices (Wan et al., 2007). In TFIDF, each utterance is considered as a *document*, and thus all utterances including the current one are regarded as *whole documents*. The frequency-based TFIDF chooses top-K words according to their TFIDF value from the set of words appearing in the meeting transcript. Since the human annotators are restricted to extract up to five keywords, the keyword extraction systems including our method are also requested to select top-5 keywords when more than five keywords are produced.

In order to see the effect of preceding utterances in baseline models, the performances are measured according to the number of preceding utterances used. Figure 3 shows the results. The X-axis of this figure is the number of preceding utterances and the Y-axis represents F-measures. As shown in this figure, the performance of the baseline models improves monotonically at first as the number of preceding utterances increases. However, the performance improvement stops when many preceding utterances are involved, and the performance begins to drop

[2]The data set is available: `http://ml.knu.ac.kr/dataset/keywordextraction.html`

[3]A guideline was given to the annotators that keywords must be a single word and the maximum number of keywords per utterance is five.

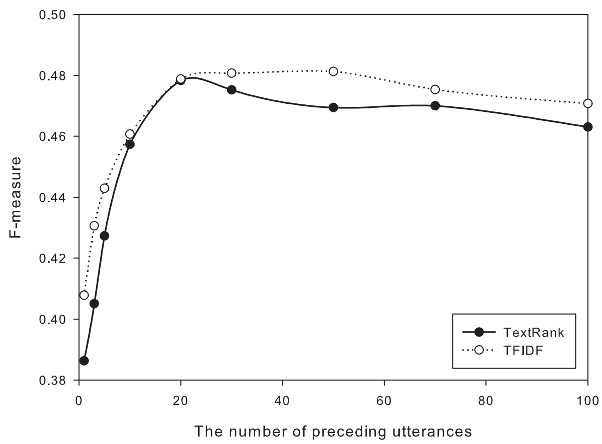

Figure 3: The performance of baseline models according to the number of preceding utterances

Table 2: The experimental results on the National Assembly transcripts

Methods	F-measure	WRS
TextRank	0.478	0.387
TFIDF	0.481	0.394
Proposed method	**0.533**	**0.421**

when too many utterances are considered. The performance of TextRank model drops from 20 preceding utterances, while that of TFIDF model begins to drops at 50 utterances. When too many preceding utterances are taken into account, it is highly possible that some of their topics are irrelevant to the current utterance, which leads to performance drop.

Table 2 compares our method with the baseline models on the National Assembly transcripts. The performances of baseline models are obtained when they show the best performance for various number of preceding utterances. TextRank model achieves F-measure of 0.478 and weighted relative score of 0.387, while TFIDF reports its best F-measure of 0.481 and weighted relative score of 0.394. Thus, the difference between TFIDF and TextRank is not significant. However, F-measure and weighted relative score of the proposed method are 0.533 and 0.421 respectively, and they are much higher than those of baseline models. In addition, our method achieves precision of 0.543 and recall of 0.523 and this is much higher performance than TextRank whose precision is just 0.510. Since the proposed method uses, as history, the preceding utterances relevant to the current utterance, its performance is kept high even if whole utterances are used. Therefore, it could be inferred that it is important to adopt only the relevant history in keyword extraction from meeting transcripts.

Table 3: The importance of temporal history

	F-measure	WRS
With Temporal History	0.533	0.421
Without Temporal History	0.518	0.413

One of the key factors of our method is the temporal history. Its importance is given in Table 3. As explained above, the temporal history is achieved by Equation (2). Thus, the proposed model does not reflect the temporal importance of preceding utterances if $w_{ij}^2 = 1$ always. That is, under $w_{ij}^2 = 1$, old utterances are regarded as important as recent utterances. Without temporal history, F-measure and weighted relative score are just 0.518 and 0.413 respectively. These poor performances prove the importance of the temporal history in keyword extraction from meeting transcripts.

5.2 ICSI Meeting Corpus in English

The proposed method is also evaluated on the ICSI meeting corpus (Janin et al., 2003) which consists of naturally occurring meetings recordings. This corpus is widely used for summarizing and extracting keywords of meetings. We followed all the experimental settings proposed by Liu et al. (2009) for this corpus. Among 26 meeting transcripts chosen by Liu et al. from 161 transcripts of the ICSI meeting corpus, 6 transcripts are used as development data and the remaining transcripts are used as data to extract keywords. The parameters for the ICSI meeting corpus are set to be $d = 0.85, W = 10$, and $\theta = 0.20$. Each meeting of the corpus consists of several topic segments, and every topic segment contains three sets of keywords that are annotated by three annotators. Up to five keywords are annotated for a topic segment.

Table 4 shows simple statistics of the ICSI meeting data. Total number of topic segments in the 26 meetings is originally 201, but some of them do not

Table 4: Simple statistics of the ICSI meeting data

Information	Value
# of meetings	26
# of topic segments	201
# of topic segments used actually	140
Average # of utterances per topic segment	260
Average # of words per utterance	7.22

Table 5: The experimental results on the ICSI corpus

Methods	F-measure	WRS
TFIDF-Liu	0.290	0.404
TextRank-Liu	0.277	0.380
ME model	0.312	0.401
Proposed method	**0.334**	**0.533**

have keywords. Such segments are discarded, and the remaining 140 topic segments are actually used. The average number of utterances in a topic segment is 260 and the average number of words per utterance is 7.22.

Unlike the National Assembly transcripts, the keywords of the ICSI meeting corpus are annotated at the topic segment level, not the utterance level. Therefore, the proposed method which extracts keywords at the utterance level can not be applied directly to this corpus. In order to obtain keywords for a topic segment with the proposed method, the keywords are first extracted from each utterance in the segment by the proposed method and then they are all accumulated. The proposed method extracts keywords for a topic segment from these accumulated utterance-level keywords as follows. Assume that a topic segment consists of l utterances. Since our method can extract up to 5 keywords for each utterance, the number of keywords for the segment can reach to $5 \cdot l$. From these keywords, we select top-5 keywords ranked by Equation (3).

The proposed method is compared with three previous studies. The first two are the methods proposed by Liu et al. (2009) One is the frequency-based method of TFIDF weighting with the features such as POS filtering, word clustering, and sentence salience score, and the other is the graph-based method with POS filtering. The last method is a maximum entropy model applied to this task (Liu et al., 2008). Note that the maximum entropy is a supervised learning model.

Table 6: The effect of considering topic relevance

Methods	F-measure	WRS
With topic relevance	0.334	0.533
Without topic relevance	0.291	0.458

Table 5 summarizes the comparison results. As shown in this table, the proposed method outperforms all previous methods. Our method achieves precision of 0.311 and recall of 0.361, and thus the F-score is 0.334. The weight relative score of the proposed method is 0.533. This is the improvement of up to 0.044 in F-measure and 0.129 in weighted relative score over other unsupervised methods (TFIDF-Liu and TextRank-Liu). It should be also noted that the proposed method outperforms even the supervised method (ME model). The difference between our method and the maximum entropy model in weighted relative score is 0.132.

One possible variant of the proposed method for the ICSI corpus is to simply merge the current utterance graph (G_1) with the history graph (G_2) rather than to extract keywords from each utterance. After the current utterance graph of the last utterance in a topic segment is merged into the history graph, the keywords for the segment are extracted from the history graph. This variant and the proposed method both rely on the temporal history, but the difference is that the history graph of the variant accumulates all information within the topic segment. Thus, the keywords extracted from the history graph by this variant are those without consideration of topic relevance.

Table 6 compares the proposed method with the variant. The performance of the variant is higher than those of TFIDF-Liu and TextRank-Liu. This proves the importance of the temporal history in keyword extraction from meeting transcripts. However, the proposed method still outperforms the variant, and it demonstrates the importance of topic relevance. Therefore, it can be concluded that the consideration of temporal history and topic relevance is critical in keyword extraction from meeting transcripts.

6 Conclusion

In this paper, we have proposed a just-in-time keyword extraction from meeting transcripts. Whenever an utterance is spoken, the proposed method extracts keywords from the utterance that best describe the utterance. Based on the graph representation of all components in a meeting, the proposed method extracts keywords by TextRank with some graph operations.

Temporal history and topic of the current utterance are two major factors especially in keyword extraction from meeting transcripts. This is because recent utterances are more important than old ones and only the preceding utterances of which topic is relevant to the current utterance are important. To model the temporal importance of the preceding utterances, the concept of forgetting curve is used in updating the history graph of preceding utterances. In addition, the subgraph of the history graph that shares words appearing in the current utterance graph is used to extract keywords rather than whole history graph. The proposed method was evaluated with the National Assembly transcripts and the ICSI meeting corpus. According to our experimental results on these data sets, the performance of keyword extraction is improved when we consider temporal history and topic relevance.

Acknowledgments

This research was supported by the Converging Research Center Program funded by the Ministry of Education, Science and Technology (2012K001342)

References

Jean Carletta. 1996. Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics*, 22(2):249–254.

Yun-Nung Chen, Yu Huang, Sheng-Yi Kong, , and Lin-Shan Lee. 2010. Automatic key term extraction from spoken course lectures using branching entropy and prosodic/semantic features. In *Proceedings of IEEE Workshop on Spoken Language Technology*, pages 265–270.

Eibe Frank, Gordon W. Paynter, Ian H. Witten, Carl Gutwin, and Craig G. Nevill-Manning. 1999. Domain-specific keyphrase extraction. In *Proceedings of the 18th International Joint Conference on Artificial intelligence*, pages 668–671.

Anette Hulth. 2003. Improved automatic keyword extraction given more linguistic knowledge. In *Proceedings of International Conference on Empirical Methods in Natural Language Processing*, pages 216–223.

Adam Janin, Don Baron, Jane Edwards, Dan Ellis, David Gelbart, Nelson Morgan, Barbara Peskin, Thilo Pfau, Elizabeth Shriberg, Andreas Stolcke, and Chuck Wooters. 2003. The icsi meeting corpus. In *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*, pages 364–367.

Fei Liu, Feifan Liu, and Yang Liu. 2008. Automatic keyword extraction for the meeting corpus using supervised approach and bigram expansion. In *Proceedings of IEEE Spoken Language Technology*, pages 181–184.

Feifan Liu, Deana Pennell, Fei Liu, and Yang Liu. 2009. Unsupervised approaches for automatic keyword extraction using meeting transcripts. In *Proceedings of Annual Conference of the North American Chapter of the ACL*, pages 620–628.

Fei Liu, Feifan Liu, and Yang Liu. 2011. A supervised framework for keyword extraction from meeting transcripts. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(3):538–548.

Rada Mihalcea and Paul Tarau. 2004. Textrank: Bringing order into texts. In *Proceedings of International Conference on Empirical Methods in Natural Language Processing*, pages 404–411.

Ani Nenkova and Rebecca Passonneau. 2004. Evaluating content selection in summarization: The pyramid method. In *Proceedings of Annual Conference of the North American Chapter of the ACL*, pages 145–152.

Peter D. Turney. 2000. Learning algorithms for keyphrase extraction. *Information Retrieval*, 2:303–336.

Peter D. Turney. 2003. Coherent keyphrase extraction via web mining. In *Proceedings of the 18th International Joint Conference on Artificial intelligence*, pages 434–439.

Xiaojun Wan and Jianguo Xiao. 2008. Collabrank: Towards a collaborative approach to single-document keyphrase extraction. In *Proceedings of International Conference on Computational Linguistics*, pages 969–976.

Xiaojun Wan, Jianwu Yang, and Jianguo Xiao. 2007. Towards an iterative reinforcement approach for simultaneous document summarization and keyword extraction. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 552–559.

Robert H. Wozniak. 1999. *Classics in Psychology, 1855–1914: Historical Essays*. Thoemmes Press.