# Fusion of Heterogeneous Information in Graph-Based Ranking for Query-Biased Summarization

Kotaro Sakamoto
Yokohama National University
sakamoto
@forest.eis.ynu.ac.jp

Hideyuki Shibuki
Yokohama National University
shib@forest.eis.ynu.ac.jp

Tatsunori Mori
Yokohama National University
mori@forest.eis.ynu.ac.jp

Noriko Kando
National Institute of Informatics
kando@nii.ac.jp

## ABSTRACT

We propose a graph-based ranking method for query-biased summarization in a three-layer graph model consisting of document, sentence and word-layers. The model has a representation that fuses three kinds of heterogeneous information: part-whole relationships between different linguistic units, similarity using the overlap of the Basic Elements (BEs) in the statements, and semantic similarity between words. In an experiment using the text summarization test collection of Nakano et al., our proposed method achieved the best result of the five considered methods, which were based on other graph models with an average R-Precision of 0.338.

## Keywords

graph-based ranking, multi-layer graph model, query-biased summarization

## 1. INTRODUCTION

Query-biased summarization, which is a multi-document summarization method customized to reflect the information need expressed in a query[10], has increased in importance for accessing user-preferred information. Following TextRank[5] and LexRank[1], which use graph-based ranking algorithms for sentence selection in summarization, several versions of graph-based ranking algorithms have been proposed for query-biased summarization[3, 4, 7, 11]. Graph-based ranking algorithms are advantageous because they do not only rely on the local context of a text unit, but rather they consider information recursively drawn from the entire text[5]. Hu et al.[3] proposed an extension of the Co-HITS-Ranking algorithm by naturally fusing sentence-level and document-level information in a graph model to take into account the strength of document-to-document and sentence-to-document correlation. Their graph model has document and sentence layers with links between two homogeneous nodes and links between two heterogeneous nodes. The homogeneous nodes are defined as nodes in the same layer, and the heterogeneous nodes are defined as ones in different layers. The link weight for homogeneous nodes is similarity based on the degree of word-overlap between two sentences or two documents, and the link weight for heterogeneous nodes is similarity based on the degree of word-overlap between a sentence and a document. Note that the link weights are homogeneous in nature (based on word overlap) even if the nodes are heterogeneous.

Here, we are interested in the behavior when link weights of different natures and different layers such as the word layer are introduced into the graph model in addition to the sentence and document layers used in the Hu et al. model. Kaneko et al.[4] proposed a four-layer graph model that consists of document, passage, sentence and word layers to comprehensively select adequate passages for summaries. In their model, two nodes from different layers are linked in accordance with part-whole relationships. For example, if a sentence contains a word, the corresponding sentence layer node is linked to the corresponding word-layer node. If another sentence contains the same word, the corresponding sentence-layer node is also linked to the same word-layer node. This is another representation of word overlap between sentences, which is distinct from word overlap using link weight. In this paper, we use a three-layer graph model, which consists of document, sentence, and word layers, based on part-whole relationships. Because we are not interested in passage selection, we do not use the passage layer. We use the Basic Elements (BEs), which are minimal semantic units and represent dependencies between the words in a sentence originally proposed by Hovy et al.[2], as units for calculating the meaning of a statement in the proposed three-layer model although Hovy et al. was not graph-based. Because BEs can more exactly represent the meaning of a statement than words, we use similarity based on the degree of BE overlap instead of word overlap as link weights in the sentence and document layers. Moreover, as link weights in the word layer, we use semantic similarity based on a thesaurus. We attempt to improve graph-based ranking by fusing the above three heterogeneous natures, which are part-whole relationships between different linguistic units, BE-overlap similarity between sentences or documents, and semantic similarity between words.

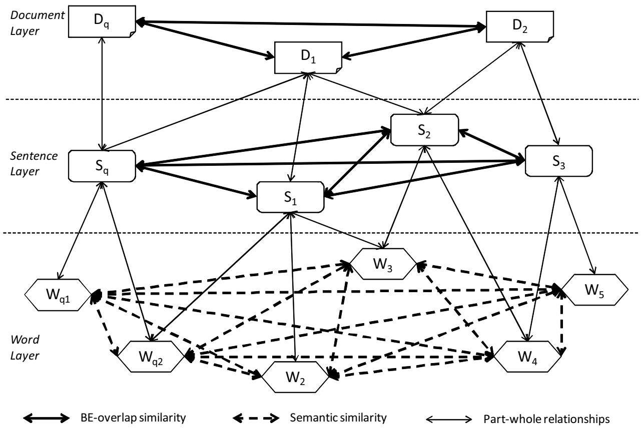


**Figure 1: The graph model for fusing heterogeneous information.**

In this paper, we propose a graph-based ranking method for query-biased summarization by extending the Co-HITS-Ranking algorithm to a three-layer graph model that has a representation fusing three kinds of heterogeneous information. Although we used Japanese texts in the experiment, the proposed graph model and algorithm are language independent. We suppose that a query is given as a sentence.

## 2. GRAPH MODEL

Figure 1 shows the graph model for fusing heterogeneous information. The model consists of three layers for representing the different linguistic units of a given document set, namely the document, sentence, and word layers.

Two nodes in the document or sentence layers are linked with each other using BE-overlap similarity. The BE-overlap similarity link is represented by a solid bold arrow in Figure 1. The BE-overlap similarity $sim_{BE}(n_i, n_j)$ between two nodes $n_i$ and $n_j$ is defined as

$$sim_{BE}(n_i, n_j) = \frac{|set_{BE}(n_i) \cap set_{BE}(n_j)|}{|set_{BE}(n_i) \cup set_{BE}(n_j)|}, \quad (1)$$

where $set_{BE}(n)$ is the set of BEs used in a linguistic unit, which is a document or sentence, corresponding to $n$. Moreover, $sim_{BE}$ $(n_i, n_j)$ is a value in the interval [0,1]. As the rate of BEs commonly used in $n_i$ and $n_j$ increases, the value of $sim_{BE}$ $(n_i, n_j)$ becomes higher.

Two nodes in the word layer are linked with each other using semantic similarity based on a thesaurus. The semantic similarity link is represented by a dashed arrow in Figure 1. The semantic similarity $sim_{sem}(n_i, n_j)$ between two word-layer nodes $n_i$ and $n_j$ is defined as

$$sim_{sem}(n_i, n_j) = \frac{MD - \max_{c \in hyper(c_k, c_l)} depth(c)}{MD}, \quad (2)$$

where $MD$ is the maximum depth of the thesaurus, $c_k$ and $c_l$ are the concepts in the thesaurus corresponding to $n_i$ and $n_j$, respectively, $hyper(c_k, c_l)$ is a set of thesaurus concepts that subsume both $c_k$ and $c_l$, and $depth(c)$ is the depth of concept $c$ in the thesaurus. Here, $sim_{sem}(n_i, n_j)$ is a value in the interval [0,1]. When the distance between nodes $n_i$ and $n_j$ decreases, the value of $sim_{sem}(n_i, n_j)$ becomes higher.

Two nodes in neighboring layers, namely between the document and sentence layers or between the sentence and word layers, are linked with each other using part-whole relationships. The part-whole relationship link is represented as a solid thin arrow in Figure 1. If a linguistic unit in the upper layer contains a unit in the lower layer, a part-whole relationship link can be drawn. For example, if a sentence in the sentence layer contains word $w$, a part-whole relationship link is drawn between the node for the sentence in the sentence layer and the node for word $w$ in the word layer. The link weight of part-whole relationships is fixed to 1. Note that a part-whole relationship link between the document and sentence layers indicates that the document contains words used in the sentence. Therefore, two nodes in the document and word layers are not directly linked.

## 3. ALGORITHM

The proposed method takes a query sentence and a set of documents as input and ranks all sentences in the documents, in order of relevance to the query, using the extended Co-HITS-Ranking algorithm. The proposed method is performed in four stages. The first stage makes a graph representing the query and documents. The second stage assigns initial ranking scores $R^q$ to all nodes in the graph. The third stage calculates homogeneous ranking scores $R^o$ according to recommendations among the neighboring homogeneous nodes. The fourth stage calculates heterogeneous ranking scores $R^e$, which are the final ranking scores, according to recommendations among the neighboring heterogeneous nodes.

### 3.1 Constructing the Graph

The graphical representation of query sentence is given as follows. The node for the query is added to the sentence layer. Another node corresponding to the query, which is regarded as a pseudo-document, is added to the document layer. Nodes of words used in the query are added to the word layer. The above-mentioned nodes are defined as query nodes in the lump. The graphical representation of the input documents is given as follows. One node per document is added to the document layer. Nodes corresponding to sentences or words used in the document are added to the sentence or word layers, respectively. Finally, two nodes in neighboring layers are linked based on part-whole relationships, two nodes in the document or sentence layers are linked using BE-overlap similarity, and two nodes in the word layer are linked using semantic similarity.

### 3.2 Assigning Initial Scores

The initial ranking score $R^q(n)$ of node $n$ is defined as

$$R^q(n) = \begin{cases} 1 & \text{(if } n \text{ is a query node)} \\ 0 & \text{(otherwise)} \end{cases} . \quad (3)$$

This is a simple criterion that $R^q(n)$ is 1 if $n$ is a query node; otherwise, $R^q(n)$ is 0.

### 3.3 Ranking Homogeneous Nodes

The ranking of homogeneous nodes in a layer is performed separately from ranking in other layers. When we define a link weight $sim^o(n_i, n_j)$ between homogeneous nodes $n_i$ and

$n_j$ as

$$sim^o(n_i, n_j) = \begin{cases} sim_{sem}(n_i, n_j) & \text{(if they are word-layer nodes)} \\ sim_{BE}(n_i, n_j) & \text{(otherwise)} \end{cases}$$

the homogeneous ranking score $R^o(n_i)$ of $n_i$ is repeatedly calculated until the value converges according to the following expression:

$$R^o(n_i) = d^o \sum_{n_j \in In(n_i)} \frac{sim^o(n_i, n_j)}{\sum_{n_k \in Out(n_j)} sim^o(n_j, n_k)} R^o(n_j) + (1 - d^o) R^q(n_i), \quad (5)$$

where $In(n_i)$ is a set of nodes linked to $n_i$, $Out(n_j)$ is a set of nodes linked from $n_j$, and $d^o$ is a trade-off parameter in the interval [0,1]. As the value of $d^o$ increases, more importance is given to ranking scores from the neighborhood homogeneous nodes compared to the initial score.

## 3.4 Ranking Heterogeneous Nodes

The ranking of heterogeneous nodes in neighboring layers is performed as follows. When a link weight $sim_{PW}(n_i, n_j)$ between heterogeneous nodes $n_i$ and $n_j$ is defined as the same value as the link weight of part-whole relationships, the heterogeneous ranking score $R^e(n_i)$ of $n_i$ is repeatedly calculated until the value converges according to the following expression:

$$R^e(n_i) = d^e \sum_{n_j \in In(n_i)} \frac{sim_{PW}(n_i, n_j)}{\sum_{n_k \in Out(n_j)} sim_{PW}(n_j, n_k)} R^e(n_j) + (1 - d^e) R^o(n_i), \quad (6)$$

where $d^e$ is a trade-off parameter in the interval [0,1]. As the value of $d^e$ increases, more importance is given to ranking scores from the neighborhood heterogeneous nodes compared to the initial score. Finally, all sentences are ranked and returned in the order of the $R^e$ values of the sentence-layer nodes, with the exception of the query node.

# 4. EXPERIMENT

## 4.1 Experimental Setup

To research effects of fusing this heterogeneous information, we perform experimental comparisons using the following four graph models. The first model has only sentence layer like TextRank or LexRank and is referred to as "Only S-layer." The second model has sentence and document layers similar to the original Co-HITS-Ranking and is referred to as "With D-layer." The third model has sentence and word layers and is referred to as "With W-layer." The forth model is the proposed model that has document, sentence and word layers and is referred to as "Three layers." Note that links of part-whole relationships are not included in the first model and that links of semantic similarity are not included in the first and second models.

For the experimental data, we use the text summarization test collection[6] annotating sentence importance as summary materials for the credibility of information on the Web. The test collection has six query sentences, six sets of Web source documents, 24 extractive summaries, and 24 free descriptive summaries. The Web source documents are retrieved via the search engine TSUBAKI[9] using query sentences. Note that the documents are already biased to a query sentence, in that they include many common words, which will influence the word- or BE-overlap similarity, such as the words used in the query sentence. All sentences in the Web documents are annotated by four human annotators with binary labels regardless of whether the sentence seems to be useful for generating the extractive summary. Note that the annotators exhaustively applied the "useful" label to sentences even if the sentences were not used as part of the extractive summary. Therefore, we evaluate ranking methods using the "useful" label. If a method can rank more "useful" sentences above "useless" sentences, the method is considered more effective than other methods.

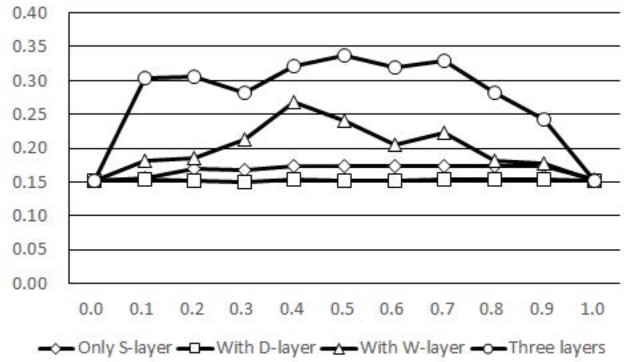


**Figure 2: Changes of the average R-Precision values .**

**Table 1: The average R-Precision values in the condition of $d^o = d^e = 0.5$.**

| | |
|---|---|
| Only S-layer | 0.173 |
| With D-layer | 0.151 |
| With W-layer | 0.240 |
| Three layers | **0.338** |

For the evaluation measure, we use the average R-Precision[1] $ARP$, which is the mean of the R-Precision values over a set of $Q$ queries. The R-Precision $RP(q)$ is the precision at the R-th position in the results ranking for query $q$ that has R "useful" sentences in the Web document set. The values $ARP$ and $RP$ are calculated as follows:

$$ARP = \frac{1}{Q} \sum_{q \in Q} RP(q), \quad (7)$$

$$RP(q) = \frac{r}{R}, \quad (8)$$

where $r$ is the number of sentences among the top R sentences that contains at least one "useful" label.

## 4.2 Result and Discussion

Figure 2 shows the changes in the average R-Precision values when the trade-off parameters $d^o$ and $d^e$ change by 0.1 between 0.0 and 1.0. Table 1 shows the average R-Precision values of the four methods at the condition that

[1]http://trec.nist.gov/pubs/trec15/appendices /CE.MEASURES06.pdf

$d^o = d^e = 0.5$. The proposed method achieved the best result. The results are improved as the number of layers in the models except for "With D-layer" increases. Therefore, we believe that fusing heterogeneous information improves the graph-based ranking algorithm and that the proposed model is effective.

Here, we describe why the result of "With D-layer" is worse than the result of "Only S-layer." The first reason is that the retrieved Web source documents are already biased to a query sentence. The second reason is that the same nature of links are used in both document and sentence layers. Therefore, the information in the document layer is very similar to the information in the sentence layer. Because the fusion of similar information cannot provide comprehensive judgment, if there is wrong information in a layer, it cannot be easily corrected by information in another layer. In the case of "With D-layer," we believe that the information of the sentence layer is deteriorated by its similar nature of the document layer. On the other hand, the proposed method was improved by fusing the word-layer information more heterogeneously than the document-layer information.

## 5. CONCLUSION

In this paper, we proposed a graph-based ranking method for query-biased summarization in a three-layer graph model that consists of document, sentence, and word layers. The model fuses part-whole relationships between different linguistic units, BE-overlap similarity between statements, and semantic similarity between words. In the experiment, the proposed method achieved the best average R-Precision of 0.338. We confirmed that fusing heterogeneous information improved the graph-based ranking algorithm when Web documents retrieved by a query sentence were given as source documents.

In our future work, we will investigate the optimal expressions for calculating the link weights and other kinds of links and layers. Moreover, we will apply this method to answer questions involving various context information. For example, at the NTCIR-11 QA-Lab task[8], a challenge to make QA systems answer questions of "world history" in real-world university entrance exams was conducted. Because such QA requires comprehensive judgment that considers various context information, we believe that the proposed method is well suited for the task.

## 6. REFERENCES


[1] G. Erkan and D. R. Radev. Lexrank: Graph-based lexical centrality as salience in text summarization. *J. Artif. Int. Res.*, 22(1):457–479, Dec. 2004.

[2] E. Hovy, C. yew Lin, and L. Zhou. A be-based multidocument summarizer with query interpretation. In *Proc. of DUC 2005*, 2005.

[3] P. Hu, D. Ji, and C. Teng. Co-hits-ranking based query-focused multi-document summarization. In *Information Retrieval Technology - 6th Asia Information Retrieval Societies Conference, AIRS 2010, Taipei, Taiwan, December 1-3, 2010. Proceedings*, pages 121–130, 2010.

[4] K. Kaneko, H. Shibuki, M. Nakano, R. Miyazaki, M. Ishioroshi, and T. Mori. Mediatory summary generation: Summary-passage extraction for information credibility on the web. In *Proceedings of the 23rd Pacific Asia Conference on Language, Information and Computation, PACLIC 23, Hong Kong, China, December 3-5, 2009*, pages 240–249, 2009.

[5] R. Mihalcea and P. Tarau. Textrank: Bringing order into texts. In D. Lin and D. Wu, editors, *Proceedings of EMNLP 2004*, pages 404–411, Barcelona, Spain, July 2004. Association for Computational Linguistics.

[6] M. Nakano, H. Shibuki, R. Miyazaki, M. Ishioroshi, K. Kaneko, and T. Mori. Construction of text summarization corpus for the credibility of information on the web. In N. C. C. Chair), K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, and D. Tapias, editors, *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta, may 2010. European Language Resources Association (ELRA).

[7] J. Otterbacher, G. Erkan, and D. R. Radev. Using random walks for question-focused sentence retrieval. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, HLT '05, pages 915–922, Stroudsburg, PA, USA, 2005. Association for Computational Linguistics.

[8] H. Shibuki, K. Sakamoto, Y. Kano, T. Mitamura, M. Ishioroshi, K. Y. Itakura, D. Wang, T. Mori, and N. Kando. Overview of the ntcir-11 qa-lab task. In *Proceedings of the 11th NTCIR Conference*, 2014.

[9] K. Shinzato, T. Shibata, D. Kawahara, and S. Kurohashi. Tsubaki: An open search engine infrastructure for developing information access methodology. *Journal of Information Processing*, 20(1):216–227, 2012.

[10] A. Tombros and M. Sanderson. Advantages of query biased summaries in information retrieval. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '98, pages 2–10, New York, NY, USA, 1998. ACM.

[11] X. Wan, J. Yang, and J. Xiao. Manifold-ranking based topic-focused multi-document summarization. In *Proceedings of the 20th International Joint Conference on Artifical Intelligence*, IJCAI'07, pages 2903–2908, San Francisco, CA, USA, 2007. Morgan Kaufmann Publishers Inc.