



Automatic Alignment between Classroom Lecture Utterances and Slide Components

Masatoshi Tsuchiya, Ryo Minamiguchi

Department of Computer Science and Engineering, Toyohashi University of Technology, Japan

{tsuchiya,minamiguchi}@is.cs.tut.ac.jp

Abstract

Multimodal alignment between classroom lecture utterances and lecture slide components is one of the crucial problems to realize a multimodal e-Learning application. This paper proposes the new method for the automatic alignment, and formulates the alignment as the integer linear programming (ILP) problem to maximize the score function which consists of three factors: the similarity score between utterances and slide components, the consistency of the explanation order, and the explanation coverage of slide components. The experimental result on the Corpus of Japanese classroom Lecture Contents (CJLC) shows that the automatic alignment information acquired by the proposed method is effective to improve the performance of the automatic extraction of important utterances.

Index Terms: Classroom lecture speech, Multimedia indexing

1. Introduction

There is increasing number of multimodal classroom lecture archives, such as the eClass project[1], the iClass project[2], the DSTC project[3], the Cornell project[4] and the Corpus of Japanese classroom Lecture Contents (CJLC) [5]. These previous researches proposed several kind of multimodal applications including a browser to listen important utterances [6]. Alignment between two modal contents (e.g. a meeting speech transcription and its meeting material) is one of the crucial problems to realize such multimodal applications [7].

This paper focuses the alignment problem to discover a lecture slide component which is explained by the specified utterance, and formulates it as the integer linear programming (ILP) problem. The conventional approach of the alignment problem is to employ the greedy algorithm which locally maximizes the similarity score function between utterances and slide components [7, 8] (Section 2). Our proposed method is different in two ways from the conventional approach: the formulation and the score function (Section 4). As already mentioned in the above, this paper formulates the alignment problem as the ILP problem, which globally maximizes the score function, instead of the greedy algorithm. Although the conventional score function considers only the similarity between utterances and slide components, the proposed function considers two additional factors to simulate good behaviors of skillful lecturers. It is entirely usual to suppose that a skillful lecturer explains his/her well-designed presentation slide in the natural order (e.g. from the top to the bottom, or from the left to the right) and that he/she explains most components and leaves few components unexplained. The first additional factor is designed for his/her natural explanation order, and the second additional factor is designed for his/her explanation coverage.

The experiment on our in-house alignment database [9], which has been constructed on the Corpus of Japanese classroom Lecture Contents (CJLC) [5], shows that those additional

factors are effective to align utterances with slide components (Section 6). The experiment on CJLC also shows that the automatic alignment information acquired by the proposed method is effective to improve the performance of the automatic extraction of important utterances (Section 5 and Section 6).

2. Related Works

There are many related works of multimodal alignment problem. This section classifies them according to two view points: distance between target contents, and criteria to extract alignment relationships.

From the view point of distance between target contents, there are two categories. The first category is that target contents are close or parallel (e.g. a speech signal and its transcription). The typical approach of the first category is to synchronize two contents with the dynamic programming (DP) algorithm based on acoustic similarity information [10]. The second category is that target contents are not parallel but are comparable at least (e.g. a meeting speech transcription and its meeting material). The typical approach of the second category is to employ the greedy algorithm to align utterances with document parts based on their similarity. This approach was examined for several types of target contents: a pair of the lecture speech transcription and its lecture slide [7], and a pair of the discussion speech transcription and its target newspaper article [8]. The other approach, a machine translation based method, was employed to align a lecture speech signal and its lecture slide, in order to improve automatic speech recognition performance [11]. This paper focuses into the alignment problem between lecture utterances and lecture slide components, thus, obviously belongs to the second category.

Alignment between two modal contents is defined as the problem to discover alignment relationships between them, according to several kind of criteria [12, 7]. The first type of criteria is thematic. The thematic alignment is considered as the problem to discover thematic relationships between target contents, like a relationship between a explanation utterance and its corresponding document part. The second type of criteria is quotation. The quotation alignment of the target contents is deterministic detection of lexicographic matching parts. The third is the reference alignment to discover referring expressions in speech transcripts and their corresponding document elements [13]. This paper focuses into the thematic alignment, to discover a slide components which is explained by a specified utterance.

3. Alignment Database

This section presents a brief summary of our in-house alignment database between classroom lecture utterances and slide components [9]. Construction procedure of the alignment database

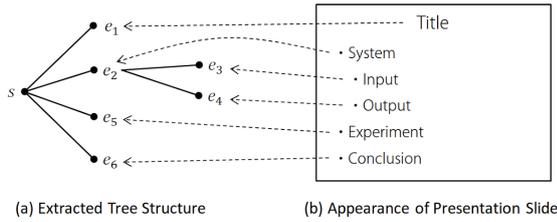


Figure 1: An example of a presentation slide. A presentation slide consists of one or more slide components.

roughly consists of two phases. The first phase is to segment target contents into alignment units, and the second phase is to align these units.

3.1. Structure of Presentation Slide

This section explains the method to decompose a presentation slide into *slide components*, which are used as alignment units of the lecture slides.

As shown in the right half of Figure 1, a presentation slide consists of one or more primitive objects, such as texts, pictures, lines and basic diagrams. These primitive objects are called as *slide components* in this paper. Slide components are carefully located in a presentation slide by its author, taking his/her presentation speech procedure into consideration. In other words, a dependency structure between slide components, represented by either their relative positional relationship or basic diagrams like an arrow sign, has strong relation to its author’s speech plan.

Because the dependency structure between slide components is not explicitly represented in the slide data itself, it is necessary to extract it. We employ the method proposed by [14], which uses relative positional relationship between slide components to extract dependency structure. Figure 1 shows an example of presentation slide designed in the conventional bulleted style and the dependency structure extracted from it. The root node represents the slide s itself. The root node has children including the headline e_1 of the slide, the first-layer bulleted text snippets e_2 , e_5 , and e_6 . And more, the node e_2 has the second-layer bulleted text snippets e_3 and e_4 as its children.

3.2. Segmentation of Utterances

This section explains the method to segment recorded classroom speech into utterances, which are used as alignment units of the lecture speeches.

There are two widely used methods to segment recorded speech signal into utterances. The first is the method which segments speech signal based on its amplitude [15, 16]. The second is the method to detect semantic boundaries on automatic speech recognition results using on various syntactic cues [17, 18]. From the view point of semantic consistency of obtained boundaries, the second method is superior to the first one, because segmentation results obtained by the first one do not match to sentence boundaries for spontaneous speech [19]. However, our target corpus, CJLC, uses the first method and its data of important sentence extraction is necessary for the latter evaluation of our proposed method, we also employ the first method in this paper, and resolve their semantic inconsistency with alignment labels.

3.3. Alignment Labels

Four labels are introduced to represent alignment relationships between utterances and slide components. First of all, Label **I** and Label **O** are introduced to distinguish whether utterances correspond to any slide components or not. Label **B** and Label **E** are introduced to resolve mismatch between automatic power-based boundary and sentence boundary described in Section 3.2. The following is more detailed descriptions of these labels.

- Label **I** means that its labeled fragment is either an utterance or a part of an utterance to explain a slide component. An explanation may be carried by either a same content word, a synonym, a hypernym, a hyponym, a paraphrase, an expression to instantiate a general case shown by the slide component into a specific case, or an expression to abstract a specific case shown by the slide component into a general case.
- Label **B** means that its labeled fragment belongs to the succeeding fragment from the view point of sentence boundary, only when the succeeding fragment has either Label **I** or Label **B**. In other words, the fragment which has Label **B** is a former part of a sentence, which must contain one or more fragments which have Label **I**.
- Label **E** is the opposite label of Label **B**, and means that its labeled fragment belongs to the preceding fragment from the view point of sentence boundary only when the preceding fragment has either Label **I** or Label **E**. In other words, the fragment which has Label **E** is a latter part of a sentence, which must contain one or more fragments which have Label **I**.
- Label **O** means that its labeled fragment are not related to any slide components.

4. Alignment between Utterances and Slide Components

This section formulates the alignment between utterances and slide components as the ILP problem to discover the best mapping matrix \hat{M} among all possible mappings from the set of utterances U to the set of slide components E .

$$\hat{M} = \operatorname{argmax}_M f(M) \quad (1)$$

A possible mapping matrix M forms $|U| \times |E|$. Its each element m_{ij} is equal to 1 if an utterance $u_i \in U$ explains a slide component $e_j \in E$, and is equal to 0 otherwise. We restrict the set of possible mapping matrices into the subset of mapping matrices which satisfies the following constraint:

$$\forall j \sum_i m_{ij} \leq 1 \quad (2)$$

This constraint means that an utterance cannot explain multiple slide components. Although a real lecture speech may not satisfy this constraint, we accept it as an approximation to ease computation complexity to discover the best mapping matrix.

The objective function $f(M)$, which represents the appropriateness of the mapping matrix M , is formulated as a linear combination of three factors as follows:

$$f(M) = \lambda_1 f_s(M) + \lambda_2 f_o(M) + \lambda_3 f_c(M), \quad (3)$$

where $f_s(M)$ is the similarity score between utterances and slide components, $f_o(M)$ is the consistency of the explanation order, and $f_c(M)$ is the explanation coverage of slide components. The parameters $\lambda_k (k = 1, 2, 3)$ are hyper parameters decided empirically through the preliminary experiments.

4.1. Similarity

The first factor $f_s(M)$ is the similarity score between utterances and slide components, which was employed in several previous researches [12, 7]. It represents a naive intuition that, when an utterance explains a slide component, their contents should be semantically similar. Thus, the first factor $f_s(M)$ is defined as the sum of individual similarity scores as follows:

$$f_s(M) = \frac{1}{|U|} \sum_i \sum_j sim(u_i, e_j) \cdot m_{ij} \quad (4)$$

Although it is possible to employ any function as the similarity function $sim(u_i, e_j)$, the standard cosine similarity function for the bag-of-words representations of u_i and e_j is employed in this paper.

4.2. Consistency of Explanation Order

The second factor $f_o(M)$, the consistency of the explanation order, aims to simulate a good behavior of a skillful lecturer. A skillful lecturer prepares well-designed presentation slides, and explains them in the natural order (e.g. from the top to the bottom, or from the left to the right). In other words, it is natural to expect that his/her explanation utterances should be ordered in the consistent way.

In order to measure the consistency of his/her explanation utterances, we introduce the *explanation distance* between two utterances. Suppose a slide component e_j which is explained by an utterance u_i , its *explanation position* $p(u_i)$ is defined as:

$$p(u_i) = j \quad (5)$$

Eq. (2) allows us to translate Eq. (5) as follows:

$$p(u_i) = \sum_j j \cdot m_{ij} \quad (6)$$

The explanation distance $d(u, u')$ between two utterances is defined as the following equation using their explanation positions:

$$d(u, u') = |p(u) - p(u')| \quad (7)$$

Because the explanation distance is reversible in explanation direction, it does not reflect the asymmetric nature of the explanation utterances: it must be rare to explain the preceding slide component after explaining the succeeding slide component. This asymmetric nature makes us to decay the explanation distance half if two utterances explain slide components in the progressing order, as follows:

$$d'(u, u') = \begin{cases} \frac{d(u, u')}{2} & \text{if } p(u) \leq p(u') \\ d(u, u') & \text{otherwise} \end{cases} \quad (8)$$

We think that the good behavior of the skillful lecturer is regarded as minimization of the sum of individual explanation distances, thus the second factor is defined as follows:

$$f_o(M) = \frac{-1}{|U| - 1} \sum_i \frac{d'(u_i, u_{i+1})}{|E|} \quad (9)$$

4.3. Coverage

The third factor $f_c(M)$, the explanation coverage of slide components, depends two assumptions. The first assumption is that a well-designed presentation slide may contain few slide components which are not explained. The second assumption is that, if a slide component is not explained, it is less important than other slide components. These two assumptions leads us to the following definition:

$$f_c(M) = \frac{1}{|E|} \sum_j rel(e_j, E) \cdot y_j \quad (10)$$

The function $rel(e_j, E)$ is the relevance score of the slide component e_j to whole topic of the set of all slide components E . The standard cosine function for bag-of-words representations of e_j and E is employed as the relevance score of the slide component. The binary variable y_j represents whether the slide component e_j is explained by one or more utterances, and satisfies the following constraints derived from the mapping matrix M :

$$y_j \in \{0, 1\}, \quad \forall i \quad y_j \geq m_{ij}, \quad y_j \leq \sum_i m_{ij}$$

5. Summarization using Alignment Information

This section describes the summarization method using alignment relationships obtained by the proposed method described in Section 4.

It is usual to formulate text summarization as the ILP problem to discover a best subset from the set of all utterances U as follows:

$$\begin{aligned} \hat{X} &= \operatorname{argmax}_X g(X) \\ &s.t. \quad \sum_i l(u_i) \cdot x_i \leq L \end{aligned} \quad (11)$$

An selection array X consists of $|U|$ binary variables. Its each element x_i is equal to 1 if an utterance u_i is selected as an important utterance, and is equal to 0 otherwise. Where $l(u)$ is the length of the utterance u , the above constraint means that the sum of lengths of the selected utterances must be shorter than the limit L .

There are two major conventional ways to define the objective function. The first is known as Maximum Marginal Relevance (MMR) [20] which was reformulated as the ILP problem [21]. The objective function of MMR is defined as follows:

$$g_m(X) = \sum_i rel(u_i, U) \cdot x_i - \sum_{i, j: i < j} red(u_i, u_j) \cdot x_i \cdot x_j \quad (12)$$

The first factor means to maximize the sum of individual semantic relevances $rel(u_i, U)$ of selected utterances, and the second factor means to minimize the sum of individual information overlaps $red(u_i, u_j)$ among selected utterances. Unfortunately, this formulation does not scale well, because the coefficient matrix of this formulation is not unimodular [22].

The second is the formulation using *conceptual units* [23, 24]. Suppose a finite set C of conceptual units which represents whole information described by the set U of all utterances. In this representation, an utterance may describe one or more conceptual units, and an information overlap between selected utterances is considered as a redundant conceptual unit(s) which

Table 1: Statistics of Alignment Database

# of slide pages	146
# of slide components	1,567
# of utterances	5,832
# of utterances related to slide components	3,237
# of Label I	2,968
# of Label B	117
# of Label E	152
# of utterances related to no slide components	2,595
# of important utterances	1,172

is described by plural utterances. The objective function of the second way is defined as follows:

$$g_c(X) = \sum_k score(c_k) \cdot z_k, \quad (13)$$

where the function $score(c_k)$ is the relevance score of the conceptual unit c_k . The binary variable z_k represents whether the conceptual unit c_k is described in the summary or not, and satisfies the following constraints:

$$z_k \in \{0, 1\}, \quad \forall i \quad z_k \geq a_{ik} \cdot x_i, \quad z_k \leq \sum_i a_{ik} \cdot x_i$$

where a_{ik} is the element of the concept explanation matrix A which forms $|U| \times |C|$. The element a_{ik} is equal to 1 if an utterance u_i describes a conceptual unit c_k , and is equal to 0 otherwise.

In this paper, the objective function is defined as a combination of both conventional ways and the coverage of slide components explained in the result summary as follows:

$$g(X) = \mu_1 \sum_i rel(u_i, U) \cdot x_i + \mu_2 \sum_j rel(e_j, E) \cdot y_j + \mu_3 \sum_k score(c_k) \cdot z_k \quad (14)$$

The binary variable y_j represents whether the slide component e_j is explained in the summary or not, and satisfies the following constraints which refer the mapping matrix M :

$$y_j \in \{0, 1\}, \quad \forall i \quad y_j \geq m_{ij} \cdot x_i, \quad y_j \leq \sum_i m_{ij} \cdot x_i$$

The parameters $\mu_k (k = 1, 2, 3)$ are hyper parameters decided empirically through the preliminary experiments. There are various possible granularity of conceptual unit, such as basic dependency subtrees obtained by trimming dependency trees [25], and weighted content words whose weights reflect their importance [24]. In this paper, the combination of the content word and the slide page is used as the conceptual unit. In other words, if the same content word is used in different two slide pages, they are counted as two conceptual units. Thus, conceptual units in this paper carry a part of slide alignment information.

6. Experiment

This section describes the evaluation experiment of our proposed alignment method and its efficiency for summarization.

Table 2 shows the evaluation experiment of our proposed alignment method, conducted against our in-house alignment database described in Section 3. The database statistics is

Table 2: Alignment results. The best accuracy is achieved when all factors are considered.

	λ_1	λ_2	λ_3	Acc.
All factors	0.1	0.6	0.3	0.547
w/o $F_s(X)$	0	0.5	0.5	0.449
w/o $F_o(X)$	0.5	0	0.5	0.498
w/o $F_c(X)$	0.5	0.5	0	0.532
Use $F_s(X)$ only	1	0	0	0.482

Table 3: Summarization results. The best performance is achieved when all factors including alignment information are considered.

	μ_1	μ_2	μ_3	ROUGE-3
All factors	0.90	0.05	0.05	0.587
w/o relevance	0	0.5	0.5	0.514
w/o slide coverage	0.5	0	0.5	0.585
w/o conceptual units	0.5	0.5	0	0.562
Use relevance only	1	0	0	0.578
Use slide coverage only	0	1	0	0.351
Use conceptual units only	0	0	1	0.515
Human	—	—	—	0.693

shown in Table 1. The chance ratio of the alignment experiment is 0.454, which is achieved when no alignment relationship is extracted from the database, because the number of utterances which explains no slide component is near to the half of the whole utterances. As shown in Table 2, our proposed methods outperform the chance ratio with an exception when the similarity score $F_s(X)$ between utterances and slide components is excluded. The best accuracy is achieved when all factors are considered, thus, the proposed factors including the consistency score $F_o(X)$ of the explanation order and the explanation coverage $F_c(X)$ are efficient to align utterances with slide components.

Table 3 shows the results of summarization experiments. The best performance is achieved when all factors including slide coverage and conceptual units are considered. Thus, this result indicates that alignment information improves the performance of automatic summarization.

7. Conclusion

This paper proposes a new multimodal alignment method for classroom lecture utterances with lecture slide components. We formulate the alignment problem as an ILP problem to maximize the score function of the possible alignment matrix. The score function includes two additional factors, the consistency score of the explanation order and the explanation coverage of slide components, in addition to the conventional similarity score between utterances and slide components. The experiment on our in-house alignment database shows that those additional factors are effective to align utterances with slide components. The experiment on CJLC also shows that the alignment information obtained by the proposed method is effective to improve the performance of the automatic extraction of important utterances.

8. Acknowledgements

This work was supported by JSPS KAKENHI Grant No. 15K12097 and No. 25280062.

9. References

- [1] J. A. Brotherton, J. R. Bhalodia, and G. D. Abowd, "Automated capture, integration, and visualization of multiple media streams," in *Proceedings. IEEE International Conference on Multimedia Computing and Systems (Cat. No.98TB100241)*, Jun 1998, pp. 54–63.
- [2] A. A. Macedo, J. A. Camacho-Guerrero, R. G. Cattelan, V. R. Inacio, Jr., and M. d. G. Campos Pimentel, "Interaction alternatives for linking everyday presentations," in *Proceedings of the Fifteenth ACM Conference on Hypertext and Hypermedia*, ser. HYPERTEXT '04. ACM, 2004, pp. 112–113. [Online]. Available: <http://doi.acm.org/10.1145/1012807.1012840>
- [3] S. Little, J. Geurts, and J. Hunter, *Dynamic Generation of Intelligent Multimedia Presentations through Semantic Inference*. Springer Berlin Heidelberg, 2002, pp. 158–175. [Online]. Available: <http://dx.doi.org/10.1007/3-540-45747-X.12>
- [4] S. Mukhopadhyay and B. Smith, "Passive capture and structuring of lectures," in *Proceedings of the Seventh ACM International Conference on Multimedia (Part 1)*, ser. MULTIMEDIA '99. ACM, 1999, pp. 477–487. [Online]. Available: <http://doi.acm.org/10.1145/319463.319690>
- [5] M. Tsuchiya, S. Kogure, H. Nishizaki, K. Ohta, and S. Nakagawa, "Developing corpus of Japanese classroom lecture speech contents," in *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, 2008.
- [6] S. Togashi and S. Nakagawa, "A browsing system for classroom lecture speech," in *Proceedings of INTERSPEECH*, 2008, pp. 2803–2806.
- [7] D. Mekhaldi, "Multimodal document alignment: towards a fully-indexed multimedia archive," in *Proceedings of multimedia information retrieval workshop (SIGIR)*, 2007.
- [8] D. Mekhaldi and D. Lalanne, "Multimodal document alignment: Feature-based validation to strengthen thematic links," *Journal of Multimedia Processing and Technologies (JMPT)*, vol. 1, no. 1, pp. 30–46, 2010.
- [9] R. Minamiguchi and M. Tsuchiya, "Developing corpus of lecture utterances aligned to slide components," in *Proceedings of the 12th Workshop on Asian Language Resources (ALR12)*. Osaka, Japan: The COLING 2016 Organizing Committee, December 2016, pp. 30–37. [Online]. Available: <http://aclweb.org/anthology/W16-5404>
- [10] K. Ohta, M. Tsuchiya, and S. Nakagawa, "Detection of precisely transcribed parts from inexact transcribed corpus," in *Proceedings of the Automatic Speech Recognition and Understanding Workshop*, Hawaii, America, December 2011, pp. 541–546.
- [11] J. Miranda, J. P. Neto, and A. W. Black, "Improving asr by integrating lecture audio and slides," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 8131–8135.
- [12] D. Mekhaldi, D. Lalanne, and R. Ingold, "Thematic alignment of recorded speech with documents," in *Proceedings of the 2003 ACM Symposium on Document Engineering*, ser. DocEng '03. New York, NY, USA: ACM, 2003, pp. 52–54. [Online]. Available: <http://doi.acm.org/10.1145/958220.958230>
- [13] A. Popescu-Belis and D. Lalanne, "Reference resolution over a restricted domain: References to documents," in *ACL 2004: Workshop on Reference Resolution and its Applications*, S. Harabagiu and D. Farwell, Eds. Barcelona, Spain: Association for Computational Linguistics, July 2004, pp. 71–78.
- [14] T. Hayama, H. Nanba, and S. Kunifuji, *PRICAI 2008: Trends in Artificial Intelligence: 10th Pacific Rim International Conference on Artificial Intelligence, Hanoi, Vietnam, December 15-19, 2008. Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, ch. Structure Extraction from Presentation Slide Information, pp. 678–687. [Online]. Available: <http://dx.doi.org/10.1007/978-3-540-89197-0.62>
- [15] N. Kitaoka, T. Yamada, S. Tsuge, C. Miyajima, T. Nishiura, M. Nakayama, Y. Denda, M. Fujimoto, K. Yamamoto, T. Takiguchi, S. Kuroiwa, K. Takeda, and S. Nakamura, "CENSREC-1-C: Development of evaluation framework for voice activity detection under noisy environment," in *IPSJ technical report, Spoken Language Processing (SIG-SLP)*, Vol. 2006, No. 107, 2006, pp. 1–6.
- [16] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-9, no. 1, pp. 62–66, 1979.
- [17] K. Shitaoka, K. Uchimoto, T. Kawahara, and H. Isahara, "Dependency structure analysis and sentence boundary detection in spontaneous Japanese," in *Proceedings of the 20th International Conference on Computational Linguistics*, ser. COLING '04. Stroudsburg, PA, USA: Association for Computational Linguistics, 2004. [Online]. Available: <http://dx.doi.org/10.3115/1220355.1220514>
- [18] Y. Akita, M. Saikou, H. Nanjo, and T. Kawahara, "Sentence boundary detection of spontaneous Japanese using statistical language model and support vector machines," in *Proceedings of INTERSPEECH*, 2006, pp. 1033–1036.
- [19] K. Takanashi, T. Maruyama, K. Uchimoto, and H. Isahara, "Identification of "sentences" in spontaneous Japanese - detection and modification of clause boundaries," in *ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition*, 2003, pp. 183–186.
- [20] J. Carbonell and J. Goldstein, "The use of MMR, diversity-based reranking for reordering documents and producing summaries," in *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 1998, pp. 335–336.
- [21] R. McDonald, *A Study of Global Inference Algorithms in Multi-document Summarization*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 557–564. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-71496-5_51
- [22] D. Roth and W.-t. Yih, "A linear programming formulation for global inference in natural language tasks," in *HLT-NAACL 2004 Workshop: Eighth Conference on Computational Natural Language Learning (CoNLL-2004)*, H. T. Ng and E. Riloff, Eds. Boston, Massachusetts, USA: Association for Computational Linguistics, May 6 - May 7 2004, pp. 1–8.
- [23] E. Filatova and V. Hatzivassiloglou, "A formal model for information selection in multi-sentence text extraction," in *Proceedings of Coling 2004*. Geneva, Switzerland: COLING, Aug 23–Aug 27 2004, pp. 397–403.
- [24] H. Takamura and M. Okumura, "Text summarization model based on maximum coverage problem and its variant," in *Proceedings of the 12th Conference of the Association for Computational Linguistics*. Association for Computational Linguistics, 2009, pp. 781–789.
- [25] E. Hovy, C.-Y. Lin, L. Zhou, and J. Fukumoto, "Automated summarization evaluation with basic elements," in *Proceedings of the Fifth Conference on Language Resources and Evaluation (LREC 2006)*, 2006, pp. 604–611.