Non-decreasing Sub-modular Function for Comprehensible Summarization

Litton J Kurisinkel Pruthwik Mishra Vigneshwaran Muralidaran
Vasudeva Varma Dipti Misra Sharma
International Institute of Information Technology, Hyderabad, India
{litton.jKurisinkel, pruthwik.mishra, vigneshwaran.m}@research.iiit.ac.in
{vv, dipti}@iiit.ac.in

Abstract

Extractive summarization techniques typically aim to maximize the information coverage of the summary with respect to the original corpus and report accuracies in ROUGE scores. Automated text summarization techniques should consider the dimensions of comprehensibility, coherence and readability. In the current work, we identify the discourse structure which provides the context for the creation of a sentence. We leverage the information from the structure to frame a *monotone (non-decreasing) sub-modular* scoring function for generating comprehensible summaries. Our approach improves the overall quality of comprehensibility of the summary in terms of human evaluation and gives sufficient content coverage with comparable ROUGE score. We also formulate a metric to measure summary comprehensibility in terms of *Contextual Independence of a sentence*. The metric is shown to be representative of human judgement of text comprehensibility.

1 Introduction

Extractive summarization techniques aim at selecting a subset of sentences from a corpus which can be a representative of the original corpus in the target summary space. Extensive work has been done on extractive summarization aimed at maximizing the information coverage of the summary with respect to the original corpus and accuracies have been reported in terms of ROUGE score. But, if a sentence is heavily dependent on its previous context in the original corpus, placing it in the summary in a different context can render a wrong inference to the reader of the summary.

The main intuition behind our approach begins with a crucial question about the linguistic nature of a text. *Is text a bag of words every time?* Psycholinguistic studies suggest that local coherence plays a vital role in inference formation while reading a text (McKoon and Ratcliff, 1992). Local coherence is undoubtedly necessary for global coherence and has received considerable attention in Computational Linguistics. ((Marcu, 2000), (Foltz et al., 1998), (Althaus et al., 2004), (Karamanis et al., 2004)). Linguistically, every sentence is uttered not in isolation but within a context in a given discourse. To make a coherent reading, *sentences use various discourse connectives that bind one sentence with another. A set of such structurally related sentences forms a Locally Coherent Discourse Unit (hereafter referred to as LDU)*. In the current work, we suggest that it is important to leverage this structural coherence to improve the comprehensibility of the generated summary. It should be noted that the concept of LDU is different from the elementary discourse units (EDUs) as discussed in Rhetorical Structure Theory(Mann and Thompson, 1988). RST is interested in describing the structure of a text in terms of relations that hold between parts of text. Any part of the text such as nuclear discourse clauses, satellite discourse clauses can be treated as elementary discourse units. In contrast, with LDUs, we are interested in identifying which sequence of sentences make up one extractable unit that has to be taken together for an extractive summarization task. The

Proceedings of NAACL-HLT 2016, pages 94–101,
San Diego, California, June 12-17, 2016.

most recent works on extractive summarization can be generalized into three steps given below:-

1. Creating an intermediate representation for the target text to capture key sentence features. The possible intermediate representations are Topic Signatures, Word-frequency count, Latent Space Approaches using Matrix Factorisations or Bayesian approaches

2. Using the intermediate representation to assign scores for individual sentence features within the text

3. Selecting a set of sentences which maximizes the total score as the summary for target text

During this process, a sentence is severed from its original context in the corpus and is eventually placed in a different context. If the level of dependence of the sentence on its context is high, then it has a higher chance to deliver an erroneous reading, when placed out of context. To understand the issue, look at the below sentences in a summary.

> *The baroque was a style of art that existed from the late 1500s to the middle of the 18th century. In 16th century, their ruler laced the old Gothic art with a newer baroque style.*

A resultant summary which contains the above two sentences one after another, can be a topically relevant summary. Both talk about 'Baroque style', 'art','century' etc and could possibly be optimal candidates for the target summary. Nevertheless, it invokes an incomprehensible reading for a human reader because the subject of the second sentence is 'their ruler' whose anaphora is not resolved in the context. Hence it is important that we do not consider a document as mere sequence of sentences or bag of words but rather as a series of LDUs.

In spite of all attempts for developing abstractive summarization techniques to mimic the human way of summarizing a text, extractive techniques still stand out as more reliable for practical purposes. So it is inevitable to enhance the extractive summarization techniques along the dimensions of readability, coherence and comprehensibility. The problem of extractive summarization can be formulated as a function maximization problem in the space of all candidate summaries as follows.

$$S^* \in argmax_{S \subseteq V} F(S) subject\ to \sum_{i \in S} c_i <= b \quad (1)$$

where F is an objective function, S^* is the summary which maximizes F with an adopted optimization method, S is a candidate summary, c_i is the cost of selecting a sentence i into summary, b is the upper bound on the total cost and V is the set of total number of sentences in the corpus.

The current work is inspired by two of the previous works namely (Lin and Bilmes, 2011) and G-Flow (Christensen et al., 2013). Lin & Bilmes observed that if the objective function to score candidate summaries is *monotone sub-modular*, a greedy approach can ensure the approximation of the summary at the global maximum by a factor of 0.632 as follows.

$$F(\hat{S}) \geq (1-1/e) * F(S_{opt}) \approx 0.632 * F(S_{opt}) \quad (2)$$

where $\hat{S}$ is the summary obtained from monotone sub-modular function F and S_{opt} is the summary at the global maximum of F.

G-Flow aimed at generating coherent summaries by constructing a sentence-level discourse graph for the entire corpus and the information from the graph is utilized to quantify the coherence of candidate summaries. In a short summary space, the sentences which structurally depend on other sentences are not encouraged. So the summaries are more comprehensible than those produced by the systems which blindly aim at achieving maximum content coverage. The need to create a discourse graph can be a big hurdle to scale the summarizer to large data sets. Also the scoring function of G-Flow is not monotone sub-modular and cannot guarantee the approximation of optimum summary as per the relation 2. *The space of current work is to establish a scheme for comprehensible summarization with a monotone sub-modular objective function*. Within the scope of this paper, when we say comprehensibility we mean **how much relevant structural context does each sentence have for better conveyability of the discourse intended by the summary**.

In the current work, we try to assign a score for each sentence based on its level of contextual independence (discussed in subsequent sections). The particular score is combined as a linear component in the candidate summary scoring function of Lin and Bilmes (Lin and Bilmes, 2011) to score sentences. While adding the third component, *monotone sub-modularity* of the scoring function is not disturbed since the contextual independence of individual sentences is constant with respect to a given corpus. We observed an improvement in system-generated summary in terms of human evaluation for comprehensibility while maintaining a reasonable level of content coverage in terms of ROUGE score.

We framed a comprehensibility index to represent the level of comprehensibility of a system generated summary using contextual independence score of individual sentences. Comprehensibility index for the generated summary is **the average contextual independence score of a sentence in the summary**. We verified, through human evaluators, whether the comprehensibility index is actually representative of the human comprehensibility.

2 Previous Work

Identification of locally coherent discourse unit (LDU) and combining the information to create a comprehensible summary is a novel problem which is not attempted by any of the previous works in the field of natural language processing to the best of our knowledge. Barzilay and Lapata(Barzilay and Lapata, 2008) attempt to measure the global coherence in terms of local coherence which is measured in terms of entity role switch while G-Flow(Christensen et al., 2013) came up with a metric to measure the coherence of the generated summary with respect to a corpus level discourse graph. Still, these two works are not directly relevant to local discourse unit identification per se.

Substantial work has been done on extractive summarization which tries to achieve a proper content coverage while reducing the redundancy. Approaches include the use of Maximum Marginal Relevance (Carbonell and Goldstein, 1998), Centroid-based Summarization (Radev et al., 2002), Summarization through Keyphrase Extraction (Qazvinian et al., 2010) and Formulation as Minimum Dominating Set problem (Shen and Li, 2010), Graph centrality to estimate the salience of a sentence (Erkan and Radev, 2004). Approaches to content analysis include generative topic models (Haghighi and Vanderwende, 2009), (Celikyilmaz and Hakkani-Tur, 2010), (Li et al., 2011b) and Discriminative models (Aker et al., 2010), ILP2 (Galanis et al., 2012) Joint Optimization of the Importance and Diversity of summary's sentences (Woodsend and Lapata, 2012), Language Model based scoring function (Takamura and Okumura, 2009) as a maximum concept coverage problem with knapsack constraint(MCKP) (Wong et al., 2008). Lin and Bilmes formulated summarization as a sub-modular function maximization problem in the possible set of candidate summaries with due respect to the summary space constraint (Lin and Bilmes, 2011).

3 Contextual Independence

Identifying whether a sentence is contextually independent or not is an important step in our approach to summarization. By Contextual Independence of a sentence, we mean that the sentence can be *globally understood even when the sentences preceding/following it are not available to the reader*. Contextual dependence, signifies only the structural dependence of a sentence in a local discourse context, not the topical dependence. Topical coherence can be captured by other parameters of optimization function used for generating summary. Take a look at the below example.

1. But it never continued after the first world war.
2. The Prime Minister of France reached Delhi yesterday.

In sentence 1, it is almost impossible to make full sense of the sentence unless the anaphor 'it' is resolved. 'But' reveals a contrast relation with the previous unmentioned sentence and therefore highly contextually dependent. Whereas a sentence like 2 can safely stand alone and convey a meaningful information even if sufficient context is not revealed. In our current work, an attempt has been made to quantify this contextual independence of a sentence in terms of surface level, generic features which are

described in subsection 4.1. Based on these features we arrived at a quantified score that denotes the probability of a sentence to be contextually independent.

4 Approach

4.1 LDU identification for measuring contextual independence

Any sentence can be identified to have a contextual dependence with another sentence based on some syntactic cues that trigger the discourse coherence. For example, a pronoun in the subject or object position of a clause in a sentence can more likely be an anaphora to a previous sentence. But extraction of such granular features and clause boundaries requires syntactic parsed output of every sentence in a document which is an overhead for the summarization system. Therefore, we have modelled the contextual independence identification of every sentence in a document as a sequence labelling problem using surface level features such as such as POS labels, unigram/bigram sequences of discourse connectives learnt across 3 windows of *W* words each. For any given sentence, we maximally take the first 3*W* words and divide them into three windows and compute the six features mentioned in Table 1 from each window.

Each of the six features signals contextual dependence. Computing these features along three windows of *W* words each is intended to statistically generalize that the features are located and computed across different clauses in a sentence. For instance, if a pronoun in one clause is resolved in the subsequent clause within the same sentence, one can safely conclude that the sentence is contextually independent. Instead of explicitly identifying the clause boundary and verifying if the anaphora is resolved within the sentence, one can generalize that if the first window does not begin with a pronoun and total number of pronouns is greater than the total number of Named Entities in the 3*W* word group, it is more likely to be resolved within the same sentence as an anaphora or cataphora. As another illustration, take for example determiners such as *the* modifying a noun as a part of prepositional phrase such as *the people from London*; the determiner 'the' in this phrase does not create any contextual dependence. This knowledge can be learnt by tracking whether the definite determiner in one window is followed by the presence of preposition in the beginning of another window. Thus the count of each of the features mentioned in Table 1 and the *W* word window boundaries are both crucial to classify a sentence as contextually dependent/independent. *W* is varied experimentally and empirically fixed as 5.

Every locally coherent discourse unit is made up of one contextually independent sentence followed by a sequence of contextually dependent sentences and hence CRF(Lafferty et al., 2001) sequence labelling algorithm is used for learning the LDUs and in turn the sequence of LDUs in an input document. The features used for contextual independence estimation are shown in Table 1.

Feature	Description
DConnect	List of commonly occurring discourse connectives
PRPcount	Count of number of pronouns
NEcount	Count of number of named entities
CC	Coordinating conjunctions
WhCount	Question words in an interrogative sentence
NounPhrase	Presence of noun phrases starting with *the*

Table 1: Feature Selection

The model predicts the probability of contextual independence of a sentence which is later used in the scoring function. The contextual dependencies include anaphors/referents, discourse connectives and determiners. The common POS tags or sequence of POS tags that signal such discourse functions are identified to be PRP, CC, DT, WP, RB, IN, TO. The reason for the choice of the features listed out in Table 1 is explained below:

DConnect and CC - Typically, a structural connection between one sentence to the next is triggered by conjunctions such as *also, nevertheless, however, but*, discourse connectives such as *for instance, in addition, according to*. These connectives usually occur at the beginning of a sentence and the features attempt to capture that in the first window.

PRPcount and NECount - Number of Pronouns and the named Entities in a 15 word group. Their relative counts together with the fact of whether they occur in initial positions of first window helps in classification

WhCount - Question words in an interrogative sentence is a marker of contextual dependency

Using the above features we are able to model the

identification of contextual independence without resorting to the overhead associated with full syntactic parsing.

4.2 Leveraging Contextual Independence Measure for Summarization

The contextual independence score of a sentence can be useful in two ways. One is to add the score as a bias term in the candidate summary scoring function and another is to exploit the same score for calculating the comprehensibility index.

4.2.1 Adding a bias term in candidate summary scoring function

Lin and Bilmes suggested a scoring function which contains weighted linear components to capture content coverage and topical diversity of the summary (Lin and Bilmes, 2011). The scoring function is given below.

$$F(S) = L_1(S) + \lambda_1 * R_1(S) \quad (3)$$

F is a monotone sub-modular function which guarantees the approximation of optimum summary by a factor of 0.632 using a greedy approach. The contextual independence of a sentence is added as a bias to the scoring function to enable the selection of contextually independent candidate sentences in the generated summary. The new scoring function is given below :

$$F(S) = L_1(S) + \lambda_1 * R_1(S) + \lambda_2 * CI(S) \quad (4)$$

Here $L_1(S)$, $R_1(S)$ and $CI(S)$ are given by equations 5, 6 and 7 respectively,

$$L_1(S) = \sum_{i \in V} min\{\sum_{j \in S} w_{i,j}, \alpha \sum_{k \in V} w_{i,k}\} \quad (5)$$

where $L_1(S)$ is the coverage function, $w_{i,j}$ is the TF-IDF cosine similarity between sentences i and j, V is the set of all sentences in the corpus, S is the set of sentences in a candidate summary, α is a learned parameter.

$$R_1(S) = \sum_{k=1}^{K} \sqrt{\sum_{j \in S \cap P_k} \frac{1}{N} \sum_{i \in V} w_{i,j}} \quad (6)$$

where $R_1(S)$ is the diversity function, N is the total no. of documents in the corpus, $P_1, P_2, ...P_k$ are sentence clusters formed out of applying k-means clustering on the set of sentences in the corpus with TF-IDF cosine similarity as the similarity metric.

$$CI(S) = \sum_{s \in S} CI(s) \quad (7)$$

where $CI(s)$ probability of a sentence s being contextually independent which is obtained from the CRF model in the section 4.1.
As per the model created in section 4.1, the contextual independence of a sentence is a constant and adding it as linear component *will not disturb the monotone sub-modularity of sentence scoring function* used by Lin and Bilmes (Lin and Bilmes, 2011)[1].

4.2.2 Framing a metric for measuring the comprehensibility of generated summary

A summary S having high CI(S) in equation 7 contains more number of Contextually Independent sentences. Therefore CI(S) represents the potential of a summary to render sufficient context for the sentences , such that the reader can grasp the same contextual interpretation from the summary sentence as is conveyed in the actual corpus, without ever reading the full corpus. *The scope of the context is captured by means of Local Discourse Unit to which the sentence belongs in the original corpus*. Instead of adding CI(S) in equation 3 directly in the scoring function, it can be utilised to frame a comprehensibility index to quantify how much a summary generated by any summarization system is comprehensible to the reader.

$$Compreh(S) = \frac{CI(S)}{N} \quad (8)$$

where Compreh(S) is the comprehensibility index, CI(S) is the contextual independence in equation 7, N is the number of sentences in the summary S.

5 Experiments and Results

We have to separately evaluate the accuracy of LDU identification, improvement of comprehensibility of system-generated summary when Contextual Independence is used as a bias term in summarization

[1] λ_1 and α take same values in Lin & Bilmes. With different trials λ_2 is empirically optimized to achieve better comprehensibility and ROUGE score and optimum value is 6

process and how much reliable the comprehensibility index is, as a metric to estimate the comprehensibility of the summary.

5.1 LDU Identification

Size	P	R	F-score	Acc%
2900	0.875	0.886	0.880	91.05

Table 2: Classification

For LDU identification model creation, we have taken a corpus containing narrative documents comprising of 2900 sentences. Two Computational Linguistics students were involved in annotation of the sentences in the corpus as either contextually dependent or independent. We obtained a Kappa score[2] of 0.703 (substantial agreement) between them. We extracted the features mentioned in Table 1 and created a training model using CRF++[3] by using 4-fold cross validation. The average precision (P), recall (R), F-score and Accuracy (Acc) were measured for different training sets and the results are shown in Table 2. The positive classification represents the contextual independence of a sentence.

5.2 CI(S) as a bias term in the scoring function

System	R	F
Nobata & Sekine	30.44	34.36
G-Flow	37.33	37.43
Best system in DUC-04	38.28	37.94
Takamura & Okumura	38.50	-
Lin & Bilmes	39.35	38.90
Our System	37.52	37.05

Table 3: ROUGE

Our System	Lin and Bilmes	Ambiguous
70%	10%	20%

Table 4: Preference of summary based on comprehensibility

By adding the CI(S) as a bias term in the scoring function in equation 3 to form 4, the system is constrained to choose the sentences which exhibit better contextual independence. Thus the equation 3 loses its flexibility in achieving maximum content coverage by the addition of CI(S) as a bias term. We have taken DUC-2004 Task2[4] dataset as our test dataset. The results for content coverage in terms of ROUGE-1 scores are given in Table 3.

The proportional decline in content coverage in terms of ROUGE score is tolerable as shown in the table 3. We have reordered the sentences in summaries generated by our system and summaries generated by Lin and Bilmes (Lin and Bilmes, 2011) implementation using the reordering system proposed by Li et al(Li et al., 2011a). Four students of Computational Linguistics participated in our evaluation experiment where we conveyed them *what we mean by comprehensibility of a summary* as defined in section 1. For each corpus in the dataset, they were made to read the documents in the corpus and asked choose the more comprehensible of the two summaries generated by our system and Lin & Bilmes provided in a random order. Our summary[5] was chosen overwhelmingly more number of times as shown in table 4.

5.3 Evaluation of Comprehensibility Index

To evaluate the comprehensibility index, we have taken into consideration, the summaries generated by Lin& Bilmes (Li et al., 2011a) and G-Flow systems(Christensen et al., 2013) for each of the corpus in DUC-2004 dataset. The four linguists participated in another evaluation experiment where we conveyed them about comprehensibility judgement like in previous experiment. For each corpus in the dataset, they were made to choose the more comprehensible of the two summaries generated by G-Flow and Lin & Bilmes provided in a random order. For the evaluation of the comprehensibility index given by equation 8, we define the accuracy of Comprehensibility Index as the percentage of times the Compreh(S) value was greater for summaries which are chosen by humans unambiguously. The details are provided in table 5. While considering both the experiments involving human evaluators, the agreement between the evaluators was 0.79 in terms of Cohen's kappa measure(Viera et al., 2005). Considering the subjective nature of annotation, we believe

[2] https://en.wikipedia.org/wiki/Fleiss'_kappa

[3] https://taku910.github.io/crfpp/

[4] http://www-nlpir.nist.gov/projects/duc/data/2004_data.html

[5] the code and annotated data are shared on https://bitbucket.org/littonj97/comprehensum/

% of times G-Flow was chosen	67%
% of times Lin & Bilmes was chosen	13%
Ambiguous	20%
Accuracy of Compreh(S)	79%
Average Compreh(S) for G-Flow	0.73
Average Compreh(S) for Lin& Bilmes	0.54

Table 5: Comprehensibility Index Evaluation Details

this is a reasonably good measure of how informative the human judgements were.

6 Future Work and Conclusion

LDU is identified currently by checking the contextual dependency of the current sentence with only the previous sentence. By using Recurrent Neural Networks this contextual dependency can be learnt beyond the preceding one sentence boundary. Comprehensibility index estimation can be improved by incorporating more information regarding topical context along with local discourse context.

References

Ahmet Aker, Trevor Cohn, and Robert Gaizauskas. 2010. Multi-document summarization using a* search and discriminative training. In *Proceedings of the 2010 conference on empirical methods in natural language processing*, pages 482–491. Association for Computational Linguistics.

Ernst Althaus, Nikiforos Karamanis, and Alexander Koller. 2004. Computing locally coherent discourses. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, page 399. Association for Computational Linguistics.

Regina Barzilay and Mirella Lapata. 2008. Modeling local coherence: An entity-based approach. *Computational Linguistics*, 34(1):1–34.

Jaime Carbonell and Jade Goldstein. 1998. 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*, pages 335–336. ACM.

Asli Celikyilmaz and Dilek Hakkani-Tur. 2010. A hybrid hierarchical model for multi-document summarization. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 815–824. Association for Computational Linguistics.

Janara Christensen, Stephen Soderland Mausam, Stephen Soderland, and Oren Etzioni. 2013. Towards coherent multi-document summarization. In *HLT-NAACL*, pages 1163–1173. Citeseer.

Günes Erkan and Dragomir R Radev. 2004. Lexrank: Graph-based lexical centrality as salience in text summarization. *Journal of Artificial Intelligence Research*, pages 457–479.

Peter W Foltz, Walter Kintsch, and Thomas K Landauer. 1998. The measurement of textual coherence with latent semantic analysis. *Discourse processes*, 25(2-3):285–307.

Dimitrios Galanis, Gerasimos Lampouras, and Ion Androutsopoulos. 2012. Extractive multi-document summarization with integer linear programming and support vector regression. In *COLING*, pages 911–926. Citeseer.

Aria Haghighi and Lucy Vanderwende. 2009. Exploring content models for multi-document summarization. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 362–370. Association for Computational Linguistics.

Nikiforos Karamanis, Massimo Poesio, Chris Mellish, and Jon Oberlander. 2004. Evaluating centering-based metrics of coherence for text structuring using a reliably annotated corpus. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, page 391. Association for Computational Linguistics.

John Lafferty, Andrew McCallum, and Fernando CN Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data.

Peifeng Li, Guangxi Deng, and Qiaoming Zhu. 2011a. Using context inference to improve sentence ordering for multi-document summarization. In *IJCNLP*, pages 1055–1061.

Peng Li, Yinglin Wang, Wei Gao, and Jing Jiang. 2011b. Generating aspect-oriented multi-document summarization with event-aspect model. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1137–1146. Association for Computational Linguistics.

Hui Lin and Jeff Bilmes. 2011. A class of submodular functions for document summarization. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 510–520. Association for Computational Linguistics.

William C Mann and Sandra A Thompson. 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text-Interdisciplinary Journal for the Study of Discourse*, 8(3):243–281.

Daniel Marcu. 2000. *The theory and practice of discourse parsing and summarization*. MIT press.

Gail McKoon and Roger Ratcliff. 1992. Inference during reading. *Psychological review*, 99(3):440.

Vahed Qazvinian, Dragomir R Radev, and Arzucan Özgür. 2010. Citation summarization through keyphrase extraction. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 895–903. Association for Computational Linguistics.

Dragomir Radev, Adam Winkel, and Michael Topper. 2002. Multi document centroid-based text summarization. In *ACL 2002*.

Chao Shen and Tao Li. 2010. Multi-document summarization via the minimum dominating set. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 984–992. Association for Computational Linguistics.

Hiroya Takamura and Manabu Okumura. 2009. Text summarization model based on maximum coverage problem and its variant. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, pages 781–789. Association for Computational Linguistics.

Anthony J Viera, Joanne M Garrett, et al. 2005. Understanding interobserver agreement: the kappa statistic. *Fam Med*, 37(5):360–363.

Kam-Fai Wong, Mingli Wu, and Wenjie Li. 2008. Extractive summarization using supervised and semi-supervised learning. In *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*, pages 985–992. Association for Computational Linguistics.

Kristian Woodsend and Mirella Lapata. 2012. Multiple aspect summarization using integer linear programming. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 233–243. Association for Computational Linguistics.