An Overview of Text Summarization Techniques

Narendra Andhale
Department of Computer Engineering
Vishwakarma Institute of Information Technology
Pune, India
narendra.andhale@gmail.com

L.A. Bewoor
Department of Computer Engineering
Vishwakarma Institute of Information Technology
Pune, India
laxmiabewoor@gmail.com

***Abstract*—Text Summarization is the process of creating a condensed form of text document which maintains significant information and general meaning of source text. Automatic text summarization becomes an important way of finding relevant information precisely in large text in a short time with little efforts. Text summarization approaches are classified into two categories: extractive and abstractive. This paper presents the comprehensive survey of both the approaches in text summarization.**

Keywords: Text summarization, extractive, abstractive.

I. Introduction

As the amount of information on the web is increasing rapidly day by day in different formats such as text, video, images. It has become difficult for individual to find relevant information of his interest. Suppose user queries for information on the internet he may get thousands of result documents which may not necessarily relevant to his concern. To find appropriate information, a user needs to search through the entire documents this causes information overload problem which leads to wastage of time and efforts. To deal with this dilemma, automatic text summarization plays a vital role. Automatic summarization condenses a source document into meaningful content which reflects main thought in the document without altering information[1]. Thus it helps user to grab the main notion within short time span. If the user gets effective summary it helps to understand document at a glance without checking it entirely, so time and efforts could be saved. Text summarization process works in three steps[2][3]:analysis, transformation and synthesis. Analysis step analyzes source text and select attributes. Transformation step transforms the result of analysis and finally representation of summary is done in synthesis step.

Text summarization approaches generally categorized into[4][5]:extractive summarization and abstractive summarization. Extractive summarization extracts important sentences or phrases from the source documents and group them to generate summary without changing the source text. However, abstractive summarization consists of understanding the source text by using the linguistic method to interpret and examine the text. The abstractive summarization aims to produce a generalized summary, conveying information in a concise way.

This paper presents extractive and abstractive text summarization techniques. The paper is organized as follows: Section II describes extractive techniques. Section III describes abstractive summarization techniques. Section IV describes limitation of both the methods. Section V concludes the paper.

II. EXTRACTIVE TEXT SUMMARIZATION METHOD

The Extractive based summarization method selects informative sentences from the document as they exactly appear in source based on specific criteria to form summary. The main challenge before extractive summarization is to decide which sentences from the input document is significant and likely to be included in the summary. For this task, sentence scoring is employed based on features of sentences[6]. It first, assigns a score to each sentence based on feature then rank sentences according to their score. Sentences with the highest score are likely to be included in final summary.

Following methods are the technique of extractive text summarization.

A. Term Frequency-Inverse Document Frequency Method

Term frequency (TF) and the inverse document frequency (IDF) are numerical statistics presents how important a word in a given document. TF is number of times a term occurs in the document and IDF is a measure that diminishes the weight of terms that occur very frequently in the collection and increases the weight of terms that occur rarely. Then sentences are scored according to product and sentence having high score are included in summary. One problem with this method is sometimes longer sentences gets high score due to fact that they contain more number of words.

Arunlfo and Ledeneva[7] proposed an approach of term selection and weighting with the help of tf-idf. They used unsupervised learning algorithm to generate non redundant summary. Sarkar[8] improved news summarization result by using sentence feature along with tf-idf. Issue regarding tf-idf is discussed in her study. Baralis et al.[9] uses weighted item set based model to accumulate information in document. This model connects various significant term then weight is given with if-idf to extract related item set to generate summary.Kamal and Sultana[10] proposed strategy depends on co-event of biological terms in sentences. Three feature terms are used to calculate the frequency of occurrence to generate summary. Jayshree and Murthy[11] calculate term frequency for extracting keywords. GSS is probabilistic feature selection when multiplied with tf-idf gives importance of word to be included in summary.

B. Cluster Based Method

Documents are composed in such a manner that they address different ideas in separate sections. It is natural to think that summaries should address different themes separated into sections of the document. In case that the document for which summary is being delivered is of entirely different subjects then summarizer assimilates this aspect through clustering. The document is represented using TF-IDF of scores of words. High frequency term represents the theme of a cluster. Summary sentence is selected based on relationship of sentence to the theme of cluster. Cluster based method generate summary of high relevance, to the given query or document topic.

Zhang and Li[12] formed a cluster of sentences using K means clustering algorithm. Based on sentence features central sentence of cluster is considered as the summary. Patil and Mahajan[13] extract and group representative sentence from a research article. Summary sentences are generated using local and global search strategy. Wu et al.[14] proposed spectral clustering and LexRank approach that leads to maximum coverage and minimum redundancy. The sparse matrix of similar sentences is generated using k-nearest neighbor method. LexRank score is calculated based on common feature to generate summary. Ferreira et al.[15] propose clustering algorithm with graph model. Document is converted into the graph, important sentences are identified using TexRank cluster is formed based on similarity between sentences. Zhang et al.[16]propose a cluster then label approach. Semi structured linked entities are clustered into semantic group and assigns a label to generate summary.

C. Text Summarization with Neural Network

A Neural Network is a processing system modeled on the human brain that tries to reenact its learning process. Neural network is an interconnected assembly of artificial neurons that uses a numerical model of computation for data processing. In case of text summarization, the strategy includes preparing the neural systems to capture the sort of sentences that ought to be incorporated into the summary. Neural Network is trained with sentences in test paragraph where each sentence is checked as to be included in summary or not. Training is done in accordance with the need of user. Neural network accurately classifies summary sentences but faces the problem of excessive training time.

Kaikhah in[17] scores each sentence according to features it contains in the feature vector. After training neural network small weight is pruned to eliminate uncommon feature. Summary is generated with high score sentences. Thu et al.[18]proposed Vietnamese text summarization based on neural network to reduce computation that uses semi-supervised learning. Sentences are scored according to word set to generate summary. Chen et al.[19]proposed recurrent neural network language model that utilize auxiliary data of word co event. Language model uses probabilistic generative paradigm to rank sentences on the basis of frequency of each unique word to generate summary. Kianmehr et al.[20]investigated the result of neural network and other summarization techniques based on feature selection. Kageback et al.[21] proposed autoencoder to derive the phrase embedding which is simple sum of words on basis of binary parse tree generated by recursive neural network. Summarization is done by measuring the similarity between phases. Prasad et al.[22] proposed part of speech disambiguation utilizing recurrent neural network .A bit vector of part of speech is given to neural network which classifies sentence to generate summary.

D. Text Summarization with Fuzzy Logic

Fuzzy Logic is a way of reasoning that resembles with the human reasoning which based on degrees of truth rather than the usual true or false (1 or 0) Boolean logic. The fuzzy system is designed with fuzzy rules and membership function which highly affect the performance. A value from zero to one is obtained for each sentence in output based on feature contained in sentence and rules defined in a knowledge base. Important sentences are extracted using IF-THEN rules based on feature criteria. Sentences are ranked in order according to score. In summary, sentence having high score are extracted. Fuzzy logic systems are simple and flexible can take imprecise, distorted, noisy input information.

Babar and Patil[23] proposed fuzzy rules and triangular membership function to score sentence based on the features. Latent semantic analysis is used to improve the summary result. Ghalehtaki et al.[24] proposed learning automata for figuring similarity of sentences, particle swarm optimization for weighting to features element. The fuzzy system classifies summary sentences based on human generated rules. Hannah et al.[25] designed fuzzy inference system for scoring sentences based on features. Rules are triggered on an aggregate score of individual sentence to generate summary. Modaresi and Conrad [26]defined a set of phrases as fuzzy set and based on membership degree of phrases, high membership values are considered as key phrases using maximum entropy model. Sentence features are used to calculate final membership value to extract summary. Suanmali et al.[27] proposed sentence extraction with selective features based on fuzzy logic. For each sentence value between zero to one is calculated based on feature and sentence is selected for summary using rules in a knowledge base.

E. Graph based Method

In this method every sentence of the document is considered as a vertex of the graph. Sentences are connected with an edge if there exist common semantic relation and based on this relation connecting edge is given weight. A graph based ranking algorithm is used to decide the importance of a vertex within a graph. Vertexes with high cardinality are considered as important sentences and included in summary. Graph based method does not require deep linguistic knowledge, nor domain knowledge for summarization. Directed graph maintains a flow of text while an edge in undirected graph captures relation using co-occurrence of term.

Mihalcea [28] investigates different ranking algorithms. In her approach similarity between sentences is determined with token overlap. Based on similarity summary is generated. Malliaros and Skianis [29] use node centrality to indicate the importance of a term in document. Local and global node centralities are considered for term weighting to form summary.Litvak and Last[30] introduced supervised and unsupervised approach for identifying the keywords based on graph syntactic presentation. In supervised algorithm is trained to identify keywords to include in summary whereas unsupervised approach uses PageRank to generate summary. Cheng et al. [31]builds dependency graph using term co-occurrence relation and syntactic relation. Using depth first traversal subgraph of three nodes is used to generate summary.

F. Latent Semantic Analysis Method

LSA is algebraic statistical method[32] that extracts meaning and resemblance of a sentence by the information about words in a particular environment. It keeps information about which words are used in sentence and reserve information of common word amongst sentences, the more common word between sentences the more it relevant. LSA extracts the source text and converts into term sentence matrix and process it through Singular Value Decomposition (SVD) for finding semantically similar words and sentences. SVD models relationships among words and sentences. The key point of LSA is it avoids the problem of synonyms but it uses only information in the input text and does not use the information of word order, syntactic relations is the major limitation of this method.

Geetha and Deepamala[33] convert document into sentence matrix where word represented by rows and sentence by column proposed by Steinberger and Jezek [34]. For each cell TF*IDF and Eagan value, Eagan vector is calculated to fill matrix cell. Sentences are selected using a cross method or average concept method to generate summary. Ozsoy et al.[35] proposed a method in which document is represented in the matrix of cells. Using the cross method sentence is selected to generate summary. They used topic method which calculates the value of concept using edge weight to extract high score sentence.

G. Machine Learning approach

In this method, the training data set is used for reference to generate summary. Summarization process is modeled as a classification problem. Sentences are classified as summary sentences and non-summary sentences based on the features that they possess. Text summarization algorithms based on machine learning approach such as Naive-Bayes, Decision Trees, Hidden Markov Model, Log-linear Models etc. are described by Das and Martins[36]. Thu and Ngoc[37] proposed Bayesian Network that finds important sentence using probability difference. Reduced sentence is generated with the weightiest path from source node to following nodes in node network. Computational cost is reduced using dynamic programming.

Sarkar et al.[38] proposed machine learning approach called bagging uses decision tree as a learner. Bag of decision trees are trained with sentence feature set. Based on decision tree training sentence get classify to include in summary. Yu and Ren[39] proposed machine learning approach that uses HMM, CRF, GMM and MMS method for cross-language summarization text. Model is trained with the feature vector. Character based tagging is used with all proposed methods to generate summary. Parkash and Shukla[40] proposed reinforcement learning where term sentence matrix for each term in the sentence is calculated using TF*IDF. For scoring the sentences, sentence signature matrix is used. Sentences are selected by calculating cosine angle of matrix reinforcement learning to generate summary.

H. Query based summarization

Query based text summarization gives right volume of the required information according to search query given by the person. Hence, the user does not need to invest extensive time for searching required information. In this summarization method the sentences in a known document are scored based on query using criteria such as frequency counts of terms. Those sentences comprising the query expressions are given higher scores than the ones containing fewer query words. Then, the sentences having maximum scores are merged into the output summary. Query based text summarization gives accurate results. If a query contains only little terms this may cause important information loss in summary.

He et al.[41] proposed feature fusion based strategy that uses similarity and skip bigram co-occurrence to calculate query relevance . User query expands with high frequency words to avoid data sparseness problem. Varadarajan and Hristidis[42] use semantic associations inside the document graph to construct query specific summary. The greedy approach is used to search query content which produces high coherent summary. Gupta and Siddiqui[43] proposed a scoring method to find the weight of sentences by using Normalized Weight. Lastly, syntactic and semantic based similarity of top sentences is measured to cluster the individual summary. Krishna et al.[44] uses statistical and linguistic techniques to calculate the relationship between the sentences in the text document and the given query. An iterative clustering algorithm is used to avoid redundancy in summary. Pandit and Potey [45] proposed a graph based approach to multi-document query specific summarization. Edge between paragraph node represents similarity determined using TF*IDF. Query words are searched over the graph to find minimum spanning tree to identify the relevant node to be included in summary.

III. ABSTRACTIVE TEXT SUMMARIZATION METHOD

Abstractive summarization generates a generalized summary by constructing new sentences alike a human being which is short and concise. Summary may contain new phrases that are not available in the source text. For generating abstractive summary language generation and compression techniques are necessary. Abstractive text summarization broadly classified into two types: Structure based and Semantic based approach.

A. Structured Based Approach

Structure based approach translates most important information from the document through cognitive schemas such as tree, ontology, lead and body phrase structure.

1) Tree Based Method: Tree based method uses dependency tree to represent text document. Source text is first represented as dependency trees then these tree are consolidating in a single tree and finally the merged dependency tree is converted to a sentence which is known as the fused sentence. The process of converting a dependency tree into a string of words is called as tree linearization. The performance of tree based summarization depends on the choice of parser and dependency preserved between words.

Barzilay and McKeown [46]uses dependency tree which fuses similar sentences using shallow parser and mapped to predicate argument structure. Sentence representing common content are determined using theme insertion algorithm. Finally, summary sentences are generated with the high rank theme using SURGE language. Filippova and Strube[47] proposed unsupervised technique which removes subtree of dependency trees to generate compressed sentences. For compression integer linear programming is used.Kikuchi et al.[48] uses word and sentence dependency to generate nested tree. Rhetorical structures and dependency parse give dependency between sentence and word. Proposed method generates a summary by trimming a nested tree using additional grammar constraints. Bing et al.[49] proposed summarization approach that uses noun , verb phrases which construct concept and fact in the original text. Phrases are extracted using dependency tree to generate constituent tree having noun and verb phrases. They used ILP to generate grammatically correct summary.

2) Template Based Method: In Template based technique template is used to represent the document .Text is matched contrary to patterns and rules to distinguish text content that mapped into template space. Template based systems differ in linguistic coverage, syntactic acquaintance, and steps involved in filling the templates. Text that fits into template indicates the content of summary. Summary generated with template based method is highly coherent. Templates are quite specific such that they accept highly relevant content and require detail semantic analysis is one of the main problem faced by template based method.

Harabagiu and Lacatusu[50] uses a template to extract information from multiple documents. Ad hoc template is iteratively filled with snippets from multiple documents that follow pattern and rules defined to generate summary. Embar et al.[51] proposed a system which uses abstraction scheme with domain template containing IR rules. Set of a template is created with variety of forms to generate summary. Oya et al.[52] use hand authored template to extract topic, important phrases from meeting transcript. Abstractive summary is generated by filling topic segment into appropriate template. Zhang et al.[53] recognize speech act to fill the template with keywords to generate template based abstractive summary. Speech act recognized with word feature, symbol based feature. Tweets are ranked based on n-gram occurrence of topic words and salient words to generate an abstractive summary.

3) Ontology Based Method: Ontology is a formal naming and definition of the entity types that are related to particular domain act as a knowledge base. In this method, a knowledge base is used to improve summarization result. Most documents on the internet are related to a particular domain having a limited vocabulary that can be better represented by the ontology. With the help of ontology attributes we can improve the semantic representation of information content and query expansion.

Lee et al. [54]presented a fuzzy system that uses ontology designed by News domain expert. Sentences are classified according to term classifier which uses ontology. The fuzzy inference mechanism calculates membership degree for each sentence according to term classifier based on domain ontology. Raghunath and Sivaranjani [55]proposed ontology based summarization that uses concept term, feature vector. They encoded ontology with tree structure node representing concept. The hierarchical classifier will select sentence according to tree structured ontology to generate summary. Hennig et al.[56] map sentences to nodes of ontology attributes. SVM classifier is trained on the binary feature. A bag of tag label is associated with each sentence represent dot product of two feature vector represents information in ontology space. Sentence of Leaf node of subtree contains specific information and classified as summary sentence. Baralis et al.[57] analyze the document to map words on yago ontology entities and calculates entity relevance score to rank sentences. Iteratively re-ranking is done to select top sentences to generate summary.

4) Lead and Body Phrase Method: This method is based on phrases. In lead and body phrase method main sentences, i.e. sentences which are informative in context and have good length are rephrase by inserting and substituting phrase. This method is good for semantically appropriate revisions for revising a lead sentence. One of the major drawbacks of Lead and body phrase is parsing degrade the performance and no generalized model for summarization.

Tanka et al.[58] search for the same chunk in lead body sentence called triggers. Phrases are identified according to similarity for substitution of body phrase into the lead phrase. This process is done iteratively to generate new summary sentences. Ishikawa et al.[59] proposed hybrid summarization method based TF method and LEAD method. Rectangular distribution function multiplied with term frequency which assigns weight to every sentence to identify importance. Sentences are ranked according to importance to compose summary. Wasson[60] investigates searchable lead sentences to summarize news document. The decision of sentence selection is based on Boolean retrieval.

5) Rule Based Method: In rule based method content selection is done with the help of information extraction rules explicitly specified by the user. Finally, language patterns are used for generating summary sentences. The strong point of this method is it creates summaries with greater information density. The main drawback is that all the rules and patterns are manually written, which is a tedious and time consuming task.

Genest and Lapalme[61] proposed abstraction scheme consist of information extraction rules by identifying noun and verb based on scheme event. The Heuristic is used to select an appropriate sentence. Patterns are designed with Simple Natural Language Generation rules to generate summary with less redundancy. Sankar and Sobha [62]find coherent chunks in the document using a set of rules and ranking text in well-organized chunks. For ranking, they use graph ranking algorithm employs word frequency, word position and string patterns to calculate the weight of sentences to generate summary. Kan and McKeown [63] proposed an approach in which machine learning employed to learn rule sets by finding occurrence patterns of predicates in manually annotated training corpora. Rules determine the content and its ordering in summary.

B. SEMANTIC BASED APPROACH

In Semantic based method, semantic representation of the document is given to natural language generation (NLG) system. This method focuses on identifying noun and verb phrases by processing linguistic data. Following are the methods of Semantic Based Approach.

1) Multimodal semantic model: In Multimodal method semantic model is constructed to capture link among concepts. The important concepts are scored based on measure and selected concept is represented as summary. The Key point of this method is coverage of information content.

Greenbacker [64] proposed approach which works in three stages first it uses an ontology to build a semantic model which represents the multimodal document. Second with information density matrix which rates a concept based on a factor such as completeness of attribute, the number of connections. Information density matrix is use to score concept and finally, summary is generated with high score concept.

2) Information item based method: In this method, a summary is generated from the abstract representation of source document. The information item is the smallest element of coherent information in a text. Information item based method provides less redundant and concise summaries.

Genest and Lapalme [65]proposed framework having information item retrieval, sentence generation, sentence selection and summary generation. In analysis part SVO triplet extracted. Sentence are generated using Simple NLG realize. The Sentence is ranked based on document frequency and summary is generated.

3) Semantic Graph Based Method: In semantic graph method, the input document is semantically represented using semantic graph. Noun and verb from the sentences are represented as graph nodes and relation between them is given by edge. It produces concise, coherent and less redundant and grammatically correct sentences.

Moawad and Aref [66] constructed a semantic graph called rich semantic graph to represent the semantic of a source document. Sentence ranking is done based on deriving the average weight of word and sentence. With highest rank sentence Rich Semantic Graph is generated and graph reduction is performed with heuristic rules to generate an abstractive summary.

IV. OBSERVATIONS OF SUMMARIZATION METHODS

The synthesized literature postulates the following major observation:

- The Main task before extractive summarization is to find important information to be comprised in the summary.
- Extracted sentences are naturally longer than average it may sometime contain unessential information in the summary.
- Significant information is present in independent sections of the document, extractive summary sometimes may not catch all informative content proliferated across the document.
- Redundant information may included in the summary.
- Extraction based summaries are unappealing to read.
- There is a lack of flow in a summary text as extracted contents which are taken from different parts of the document leads to sudden topic shift.
- Summary representation is the biggest problem of abstractive method.
- Sometimes abstractive summary does not capture the semantic relationship between important terms in the document.
- For producing generalized summary Natural Language Generation rules are highly needed.
- Abstractive summaries are sometimes results in incoherent.
- Abstraction requires semantic understanding of text.
- The quality of abstractive summary depends on the deep linguistic skills.

V. CONCLUSION

This survey paper covers extractive and abstractive summarization techniques. Summarization system should produce an effective summary in a short time with less redundancy having grammatically correct sentences. Both extractive and abstractive method yields good result according to the context in which they used. The reviewed literature opens up the challenging area for hybridization of these methods to produce informative, well compressed and readable summaries.

REFERENCES

[1] E. Lloret and M. Palomar, "Text summarisation in progress : a literature review," in *Springer*, no. April 2011, pp. 1–41, Springer, 2012.

[2] U. Hahn and I. Mani, “of Automatic Researchers are investigating summarization tools and methods that,” in *IEEE Computer 33.11*, no. November, pp. 29–36, IEEE, 2000.

[3] K. Spärck Jones, “Automatic summarising: The state of the art,” *Information Processing & Management*, vol. 43, pp. 1449–1481, nov 2007.

[4] V. Gupta and G. S. Lehal, “A Survey of Text Summarization Extractive techniques,” *Journal of Emerging Technologies in Web Intelligence*, vol. 2, no. 3, pp. 258–268, 2010.

[5] A. Khan and N. Salim, “A review on abstractive summarization methods,” *Journal of Theoretical and Applied Information Technology*, vol. 59, no. 1, pp. 64–72, 2014.

[6] R. Ferreira, L. De Souza Cabral, R. D. Lins, G. Pereira E Silva, F. Freitas, G. D. C. Cavalcanti, R. Lima, S. J. Simske, and L. Favaro, “Assessing sentence scoring techniques for extractive text summarization,” *Expert Systems with Applications*, vol. 40, no. 14, pp. 5755–5764, 2013.

[7] R. A. García-Hernández and Y. Ledeneva, “Word sequence models for single text summarization,” in *Proceedings of the 2nd International Conferences on Advances in Computer-Human Interactions, ACHI 2009*, pp. 44–48, IEEE, 2009.

[8] K. Sarkar, “An Approach to Summarizing Bengali News Documents,” in *International Conference on Advances in Computing, Communications and Informatics*, pp. 857–862, ACM, 2012.

[9] E. Baralis, L. Cagliero, A. Fiori, and P. Garza, “MWI-Sum: A Multilingual Summarizer Based on Frequent Weighted Itemsets,” *ACM Transactions on Information Systems*, vol. 34, no. 1, pp. 1–35, 2015.

[10] M. Kamal, “Text Summarization in Bioinformatics,” in *15th International Conference on Computer and Information Technology*, pp. 592–597, IEEE, 2012.

[11] R. Jayashree, Srikanta Murthy K, and B. Anami, “Categorized Text Document Summarization in the Kannada Language by sentence ranking,” in *2012 12th International Conference on Intelligent Systems Design and Applications (ISDA)*, pp. 776–781, IEEE, nov 2012.

[12] P.-y. Zhang and C.-h. Li, “Automatic text summarization based on sentences clustering and extraction,” in *2nd IEEE International Conference on Computer Science and Information Technology*, vol. 1, pp. 167–170, IEEE, 2009.

[13] S. R. Patil and S. M. Mahajan, “A novel approach for research paper abstracts summarization using cluster based sentence extraction,” in *Proceedings of the International Conference & Workshop on Emerging Trends in Technology - ICWET ’11*, no. Icwet, p. 583, ACM Press, 2011.

[14] K. Wu, P. Shi, and D. Pan, “An Approach to Automatic Summarization For Chinese Text Based on the Combination of Spectral Clustering and LexRank,” in *12th International Conference on Fuzzy Systems and Knowledge Discovery*, pp. 1350–1354, IEEE, 2015.

[15] R. Ferreira, L. de Souza Cabral, F. Freitas, R. D. Lins, G. de França Silva, S. J. Simske, and L. Favaro, “A multi-document summarization system based on statistics and linguistic treatment,” *Expert Systems with Applications*, vol. 41, pp. 5780–5787, oct 2014.

[16] K. Zhang, Y. Xiao, H. Tong, H. Wang, and W. Wang, “WiiCluster: a Platform for Wikipedia Infobox Generation,” in *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management - CIKM ’14*, pp. 2033–2035, ACM Press, 2014.

[17] K. Kaikhah, “Automatic text summarization with neural networks,” in *2nd International Conference on ’Intelligent Systems’.*, no. June, pp. 40–44, IEEE, 2004.

[18] Ha Nguyen Thi Thu, Quynh Nguyen Huu, and Tu Nguyen Thi Ngoc, “A supervised learning method combine with dimensionality reduction in Vietnamese text summarization,” in *Computing, Communications and IT Applications Conference (ComComAp)*, vol. 9,12,,4903, pp. 69–73, IEEE, apr 2013.

[19] K.-Y. Chen, S.-H. Liu, B. Chen, H.-M. Wang, E.-E. Jan, W.-L. Hsu, and H.-H. Chen, “Extractive Broadcast News Summarization Leveraging Recurrent Neural Network Language Modeling Techniques,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, pp. 1322–1334, aug 2015.

[20] K. Kianmehr, S. Gao, J. Attari, and M. M. Rahman, “Text Summarization Techniques : SVM versus Neural Networks ,” in *International Conference on Information Integration and Web-based Applications & Services*, pp. 487–491, ACM, dec 2009.

[21] M. Kageback, O. Mogren, N. Tahmasebi, and D. Dubhashi, “Extractive Summarization using Continuous Vector Space Models,” in *Proceedings of the 2nd Workshop on Continuous Vector Space Models and their Compositionality (CVSC)*, pp. 31–39, Association for Computational Linguistics, apr 2014.

[22] R. S. Prasad, U. Kulkarni, and J. R. Prasad, “A novel evolutionary connectionist text summarizer (ECTS),” in *3rd International Conference on Anti-counterfeiting, Security, and Identification in Communication*, pp. 606–610, IEEE, aug 2009.

[23] S. Babar and P. D. Patil, “Improving Performance of Text Summarization,” *Procedia Computer Science*, vol. 46, no. Icict 2014, pp. 354–363, 2015.

[24] R. A. Ghalehtaki, H. Khotanlou, and M. Esmaeilpour, “A combinational method of fuzzy, particle swarm optimization and cellular learning automata for text summarization,” in *Iranian Conference on Intelligent Systems (ICIS)*, vol. 15, pp. 1–6, IEEE, feb 2014.

[25] M. E. Hannah, T. V. Geetha, and S. Mukherjee, “Automatic Extractive Text Summarization Based on Fuzzy Logic: A Sentence Oriented Approach,” in *Swarm Evolutionary and Memetic Computing Conference*, vol. 7076, pp. 530–538, Springer, 2011.

[26] P. Modaresi and S. Conrad, “From Phrases to Keyphrases:An Unsupervised Fuzzy Set Approach to Summarize News Articles,” in *Proceedings of the 12th International Conference on Advances in Mobile Computing and Multimedia - MoMM ’14*, pp. 336–341, ACM Press, dec 2014.

[27] L. Suanmali, M. S. Binwahlan, and N. Salim, “Sentence Features Fusion for Text Summarization Using Fuzzy Logic,” in *2009 Ninth International Conference on Hybrid Intelligent Systems*, vol. 1, pp. 142–146, IEEE, 2009.

[28] R. Mihalcea, “Graph-based ranking algorithms for sentence extraction, applied to text summarization,” in *Proceedings of the ACL 2004 on Interactive poster and demonstration sessions ACLdemo 04*, no. 20, pp. 20–es, Association for Computational Linguistics, 2004.

[29] F. D. Malliaros and K. Skianis, “Graph-Based Term Weighting for Text Categorization,” in *Proceedings of International Conference on Advances in Social Networks Analysis and Mining - ASONAM ’15*, pp. 1473–1479, ACM Press, aug 2015.

[30] M. Litvak and M. Last, “Graph-based keyword extraction for single-document summarization,” in *Proceedings of the Workshop on Multi-source Multilingual Information Extraction and Summarization - MMIES ’08*, pp. 17–24, Association for Computational Linguistics, aug 2008.

[31] K. Cheng, Y. Li, and X. Wang, “Single Document Summarization Based on Triangle Analysis of Dependency Graphs,” in *16th International Conference on Network-Based Information Systems*, pp. 38–43, IEEE, sep 2013.

[32] T. K. Landauer, P. W. Foltz, and D. Laham, “An introduction to latent semantic analysis,” *Discourse Processes*, vol. 25, pp. 259–284, jan 1998.

[33] Geetha J K and Deepamala N, “Kannada text summarization using Latent Semantic Analysis,” in *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 1508–1512, IEEE, aug 2015.

[34] J. Steinberger and K. Jezek, “Using latent semantic analysis in text summarization and summary evaluation,” in *7th International Conference ISIM*, no. April, pp. 93–100, 2004.

[35] M. G. Ozsoy, F. N. Alpaslan, and I. Cicekli, “Text summarization using Latent Semantic Analysis,” in *Proceedings of the 23rd International Conference on Computational Linguistics*, vol. 37, pp. 405–417, Coling, aug 2011.

[36] D. Das and A. F. T. Martins, “A Survey on Automatic Text Summarization,” *Carnegie Mellon University*, pp. 1–31, 2007.

[37] H. N. T. Thu and D. V. T. Ngoc, “Improve Bayesian Network to Generating Vietnamese Sentence Reduction,” in *International Conference on Future Information Engineering*, vol. 10, pp. 190–195, Elsevier B.V., 2014.

[38] K. Sarkar, M. Nasipuri, and S. Ghose, “Using Machine Learning for Medical Document Summarization,” *International Journal of Database Theory and Application*, vol. 4, no. 1, pp. 31–48, 2011.

[39] L. Yu and F. Ren, “A study on cross-language text summarization using supervised methods,” in *International Conference on Natural Language Processing and Knowledge Engineering*, pp. 1–7, IEEE, sep 2009.

[40] C. Prakash and A. Shukla, “ Human Aided Text Summarizer SAAR using Reinforcement Learning,” in *International Conference on Soft Computing and Machine Intelligence*, vol. 474010, pp. 83–87, IEEE, sep 2014.

[41] T. He, F. Li, W. Shao, J. Chen, and L. Ma, “A New Feature-Fusion Sentence Selecting Strategy for Query-Focused Multi-document

Summarization," in *International Conference on Advanced Language Processing and Web Information Technology*, pp. 81–86, IEEE, 2008.

[42] R. Varadarajan and V. Hristidis, "A system for query-specific document summarization," in *Proceedings of the 15th international conference on Information and knowledge management - CIKM '06*, p. 622, ACM Press, nov 2006.

[43] V. K. Gupta and T. J. Siddiqui, "Multi-document summarization using sentence clustering," in *4th International Conference on Intelligent Human Computer Interaction*, pp. 1–5, IEEE, dec 2012.

[44] R. V. V. M. Krishna, S. Y. P. Kumar, and C. S. Reddy, "A hybrid method for query based automatic summarization system," *International Journal of Computer Applications*, vol. 68, no. 6, pp. 39–43, 2013.

[45] S. R. Pandit and M. Potey, "A Query Specific Graph Based Approach to Multi-document Text Summarization: Simultaneous Cluster and Sentence Ranking," in *International Conference on Machine Intelligence and Research Advancement*, pp. 213–217, IEEE, dec 2013.

[46] R. Barzilay and K. R. McKeown, "Sentence Fusion for Multidocument News Summarization," *Computational Linguistics*, vol. 31, pp. 297–328, sep 2005.

[47] K. Filippova and M. Strube, "Dependency tree based sentence compression," in *Fifth International Natural Language Generation Conference on - INLG '08*, p. 25, Association for Computational Linguistics, 2008.

[48] Y. Kikuchi, T. Hirao, H. Takamura, M. Okumura, and M. Nagata, "Single Document Summarization based on Nested Tree Structure," in *52nd Annual Meeting of the Association for Computational Linguistics*, pp. 315–320, ACL, jun 2014.

[49] Y. L. W. L. Lidong Bing, Piji Li and P. R. J., Weiwei Guo, "Abstractive Multi-Document Summarization via Phrase Selection and Merging," in *7th International Joint Conference on Natural Language Processing*, pp. 1587–1597, Association for Computational Linguistics, dec 2015.

[50] S. Harabagiu and F. Lacatusu, "Generating single and multi-document summaries with GISTexter," in *Document Understanding Conference 2002 (DUC'02)*, pp. 30–38, 2002.

[51] V. R. Embar, S. R. Deshpande, A. K. Vaishnavi, V. Jain, and J. S. Kallimani, "sArAmsha - A Kannada abstractive summarizer," in *International Conference on Advances in Computing, Communications and Informatics*, pp. 540–544, IEEE, aug 2013.

[52] T. Oya, Y. Mehdad, G. Carenini, and R. Ng, "A Template-based Abstractive Meeting Summarization : Leveraging Summary and Source Text Relationships," in *8th International Natural Language Generation Conference*, no. June, pp. 45–53, Association for Computational Linguistics, 2014.

[53] Renxian Zhang, Wenjie Li, Dehong Gao, and You Ouyang, "Automatic Twitter Topic Summarization With Speech Acts," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, pp. 649–658, mar 2013.

[54] C. Lee, Z. Jian, and L. Huang, "A Fuzzy Ontology and Its Application to News Summarization," *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, vol. 35, pp. 859–880, oct 2005.

[55] R. Ragunath and N. Sivaranjani, "Ontology Based Text Document Summarization System Using Concept Terms," *ARPN Journal of Engineering and Applied Sciences*, vol. 10, no. 6, pp. 2638–2642, 2015.

[56] L. Hennig, W. Umbrath, and R. Wetzker, "An Ontology-Based Approach to Text Summarization," in *International Conference on Web Intelligence and Intelligent Agent Technology*, vol. 3, pp. 291–294, IEEE, dec 2008.

[57] E. Baralis, L. Cagliero, S. Jabeen, A. Fiori, and S. Shah, "Multi-document summarization based on the Yago ontology," *Expert Systems with Applications*, vol. 40, pp. 6976–6984, dec 2013.

[58] H. Tanaka, A. Kinoshita, T. Kobayakawa, T. Kumano, and N. Kato, "Syntax-driven sentence revision for broadcast news summarization," in *Proceedings of the Workshop on Language Generation and Summarisation - UCNLG+Sum '09*, no. August, pp. 39–47, Association for Computational Linguistics, aug 2009.

[59] K. Ishikawa, S. Ando, and A. Okumura, "Hybrid Text Summarization Method based on the TF Method and the LEAD Method," in *In Proceedings of the 2nd National Institute of Informatics Test Collection Information Retrieval (NTCIR) Workshop*, no. 1, pp. 5–219, 2001.

[60] M. Wasson, "Using leading text for news summaries:Evaluation Results and Implications for Commercial Summarization Applications," in *17th international conference on Computational linguistics*, vol. 2, pp. 1364–1368, Association for Computational Linguistics, 1998.

[61] P.-E. Genest and G. Lapalme, "Fully Abstractive Approach to Guided Summarization," *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistic*, pp. 354–358, jul 2012.

[62] K. Sankar and L. Sobha, "An approach to text summarization," in *Proceedings of Third International Cross Lingual Information Access Workshop- CLIAWS3*, pp. 53–60, Association for Computational Linguistics, jun 2009.

[63] M.-Y. Kan and K. R. McKeown, "Corpus-trained text generation for summarization.," in *Second International Natural Language Generation Conference*, vol. 20, pp. 359–392, Association for Computational Linguistics, 2002.

[64] C. Greenbacker, "Towards a framework for abstractive summarization of multimodal documents," in *49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pp. 75–80, ACL, jun 2011.

[65] P. E. Genest and G. Lapalme, "Framework for abstractive summarization using text-to-text generation," in *49th Annual Meeting of the Association for Computational Linguistics*, no. June, pp. 64–73, Association for Computational Linguistics, jun 2011.

[66] I. F. Moawad and M. Aref, "Semantic graph reduction approach for abstractive Text Summarization," in *Seventh International Conference on Computer Engineering & Systems (ICCES)*, pp. 132–138, IEEE, nov 2012.