CrossMark

Recent automatic text summarization techniques: a survey

Mahak Gambhir[1] · **Vishal Gupta**[1]

Abstract As information is available in abundance for every topic on internet, condensing the important information in the form of summary would benefit a number of users. Hence, there is growing interest among the research community for developing new approaches to automatically summarize the text. Automatic text summarization system generates a summary, i.e. short length text that includes all the important information of the document. Since the advent of text summarization in 1950s, researchers have been trying to improve techniques for generating summaries so that machine generated summary matches with the human made summary. Summary can be generated through extractive as well as abstractive methods. Abstractive methods are highly complex as they need extensive natural language processing. Therefore, research community is focusing more on extractive summaries, trying to achieve more coherent and meaningful summaries. During a decade, several extractive approaches have been developed for automatic summary generation that implements a number of machine learning and optimization techniques. This paper presents a comprehensive survey of recent text summarization extractive approaches developed in the last decade. Their needs are identified and their advantages and disadvantages are listed in a comparative manner. A few abstractive and multilingual text summarization approaches are also covered. Summary evaluation is another challenging issue in this research field. Therefore, intrinsic as well as extrinsic both the methods of summary evaluation are described in detail along with text summarization evaluation conferences and workshops. Furthermore, evaluation results of extractive summarization approaches are presented on some shared DUC datasets. Finally this paper concludes with the discussion of useful future directions that can help researchers to identify areas where further research is needed.

✉ Vishal Gupta
  vishal_gupta100@yahoo.co.in

  Mahak Gambhir
  gambhir.mahak@gmail.com

[1] University Institute of Engineering and Technology, Panjab University, Chandigarh, India

Springer

1 Introduction

Automatic text summarization system generates summary, i.e., condensed form of the document that contains a few important sentences selected from the document. In late fifties (Luhn 1958), text summarization began and till now, there is great improvement in this field. A large number of techniques and approaches have been developed in this field of research (Jones 2007). A summary generated by an automatic text summarizer should consist of the most relevant information in a document and at the same time, it should occupy less space than the original document. Nevertheless, automatic summary generation is a challenging task. There are many issues like redundancy, temporal dimension, co-reference, sentence ordering, etc that need particular attention when summarizing multiple number of documents, thereby, making this task more complex (Goldstein et al. 2000).

1.1 Need of text summarization

With the increase in on-line publishing, large internet users and fast development of electronic government (e-government), need of text summarization has emerged. As the information-communication technologies are growing at a great speed, a large number of electronic documents are available on-line and users are facing difficulty to find relevant information. Moreover, internet has provided large collections of text on a variety of topics. This accounts for the redundancy in the texts available on-line. Users get so exhausted reading large amount of texts that they may skip reading many important and interesting documents. Therefore, robust text summarization system is currently needed in this generation. These systems can compress information from various documents into a shorter length, readable summary (Yang and Wang 2008; Harabagiu and Lacatusu 2010). Four main objectives are considered by Huang et al. (2010): coverage of information, information significance, redundancy in information and cohesion in text.

2 Various types of text Summarization

On the basis of number of documents, single and multi-document summarizations are the two important categories of summarization (Zajic et al. 2008; Fattah and Ren 2009). Summary is generated from a single document in single document summarization whereas in multi-document summarization, many documents are used for generating a summary. It is considered that summarization of single document is extended to generate summarization of multiple documents. But the task of summarizing multiple documents is more difficult than the task of summarizing single documents. Redundancy is one of the biggest problems in summarizing multiple documents. There are some systems that tackle with redundancy by initially selecting the sentences at the beginning of the paragraph and then measuring the similarity of the next sentence with the already chosen sentences and if this sentence consists of relevant new content, then only it is selected (Sarkar 2010). MMR approach (Maximal Marginal Relevance) is suggested by Carbonell and Goldstein (1998) for reducing redundancy. Researchers from all over the world are investigating different methods to produce

best results in multi-document summarization (Tao et al. 2008; Wan 2008; Wang et al. 2011, 2008a, b, 2009).

Extractive or Abstractive summarization is also a classification of document summarization. An extract summary is generated in extractive summarization by selecting a few relevant sentences from the original document. Summary's length depends on the compression rate. It is a simple and robust method for summarization of text. Here some saliency scores are assigned to sentences in the documents and then highly scored sentences are chosen to generate the summary. Whereas abstractive summarization produces an abstract summary which includes words and phrases different from the ones occurring in the source document. Therefore, abstract is a summary that consists of ideas or concepts taken from the original document but are re-interpreted and shown in a different form. It needs extensive natural language processing. Therefore, it is much more complex than extractive summarization. Thus, extractive summarization, due its increased feasibility has attained a standard in summarization of documents.

Summaries can also be of two types: generic or query-focused (Gong and Liu 2001; Dunlavy et al. 2007; Wan 2008; Ouyang et al. 2011). Topic-focused or user-focused summaries are the other names for query-focused summaries. Such a summary includes the query related content whereas a general sense of the information present in the document is provided in a generic summary.

Summarization task can be either supervised or unsupervised (Mani and Maybury 1999; Fattah and Ren 2009; Riedhammer et al. 2010). Training data is needed in a supervised system for selecting important content from the documents. Large amount of labeled or annotated data is needed for learning techniques. These systems are addressed at sentence level as two-class classification problem in which sentences belonging to the summary are termed as positive samples and sentences not present in the summary are named as negative samples (Song et al. 2011; Chali and Hasan 2012). For performing sentence classification, some popular classification methods are employed such as Support Vector Machine (SVM) (Ouyang et al. 2011) and neural networks (Fattah and Ren 2009). On the other hand, unsupervised systems do not require any training data. They generate the summary by accessing only the target documents. Thus, they are suitable for any newly observed data without any advanced modifications. Such systems apply heuristic rules to extract highly relevant sentences and generate a summary (Fattah and Ren 2009). The technique employed in unsupervised system is clustering.

Based on the style of output, there are two types of summaries: indicative and informative summaries. Indicative summaries tell what the document is about. They give information about the topic of the document. Informative summaries, while covering the topics, give the whole information in elaborated form.

There exists one more summary similar to them, critical evaluation abstracts. These summaries consist of views of author about a particular topic and contain opinions, reviews, recommendations, feedbacks, etc. For e.g., reviewers review the research paper for the journals and conferences and send back their valuable feedback to the candidate that includes either acceptance, rejection or acceptance of the paper with some modifications.

On the basis of language, there are three kinds of summaries: multi-lingual, mono-lingual and cross-lingual summaries. When language of source and target document is same, it's a mono-lingual summarization system. When source document is in a number of languages like English, Hindi, Punjabi and summary is also generated in these languages, then it is termed as a multi-lingual summarization system. If source document is in English and the summary generated is in Hindi or any other language other than English, then it is known as a cross-lingual summarization system.

Another common type is Web-based summarization. Nowadays users are facing information in abundance on the internet. Web pages on internet are doubling every year. Some search engines like Google Fast, Alta Vista, etc help users to find the information they require but they return a list of large number of web pages for a single query. As a result, users need to go through multiple pages to know which documents are relevant and which are not and most of the users give up their search in the first try. Therefore, web based summaries summarize important information present in the web pages. Radev et al. (2001) proposed WebInEssence, an effective search engine that can summarize clusters of related documents which can help users to explore retrieval results systematically.

E-mail based summarization is also a type of summarization in which email conversations are summarized. Email has become an effective way of communication because of its high delivery speed and lack of cost. Emails keep on coming in the inbox due to which email overloading problem occurs and large time is spent in reading, sorting and archiving the incoming emails. There are also some other uses of email summaries. In the world of business, email summarization can be used as a corporate memory where thread summaries convey all business decisions made in the past.

Personalized summaries contain the specific information that the user desires. Different consumers have different requirements so such systems after determining the user's profile select the important content for generating the summary. In Update summaries, it is considered that consumers have the basic information about the topic and requires only the current updates regarding the topic.

Web 2.0 has caused the development of some new kinds of websites like social networking sites, forums, blogs, etc where users depict their feelings or give reviews on a product, entity, service or topic. This has led to the emergence of sentiment-based summaries. Text Summarization (TS) and Sentiment Analysis (SA) together form opinion mining and they work together for generating such summaries. In such summaries opinions are initially detected and classified on the basis of subjectivity (whether the sentence is subjective or objective) and then on the basis of polarity (positive, negative or neutral) (Pang and Lee 2008).

Survey summaries obtain a normal overview of a specific topic or entity. These are usually lengthy as they contain most important facts, regarding person, place or any other entity. Survey summaries, biographical summaries and Wikipedia articles, all come under this category. Table 1 below describes different types of summaries along with the factors determining the summarization type.

3 Classification of extractive approaches for summary generation

3.1 Statistical based approaches

These approaches deal with some statistical features which help to extract important sentences and words from source text. These techniques are independent of any language, such that if the summarizer is developed using these techniques, then it can summarize text in any language. So, these techniques do not require any additional linguistic knowledge or complex linguistic processing (Ko and Seo 2008). Also, they require less processor and memory capacity. Some of the statistical features (Fattah and Ren 2009) are position of sentence, positive keyword (based on frequency count), negative keyword (based on frequency count), centrality of sentence (i.e. similarity with other sentences), resemblance of sentence to the title, relative length of the sentence, presence of numerical data in the sentence, presence of proper noun

Table 1 Different types of summaries on the basis of various factors

S. No.	Types of summary	Factors
1.	Single and multi-document	Number of documents
2.	Extractive and abstractive	Output (if extract or abstract is required)
3.	Generic and query-focused	Purpose (whether general or query related data is required)
4.	Supervised and unsupervised	Availability of training data
6.	Mono, multi and cross-lingual	Language
7.	Web-based	For summarizing web pages
8.	E-mail based	For summarizing e-mails
9.	Personalized	Information specific to a user's need
10.	Update	Current updates regarding a topic
11.	Sentiment-based	Opinions are detected
12.	Survey	Important facts regarding person, place or any other entity
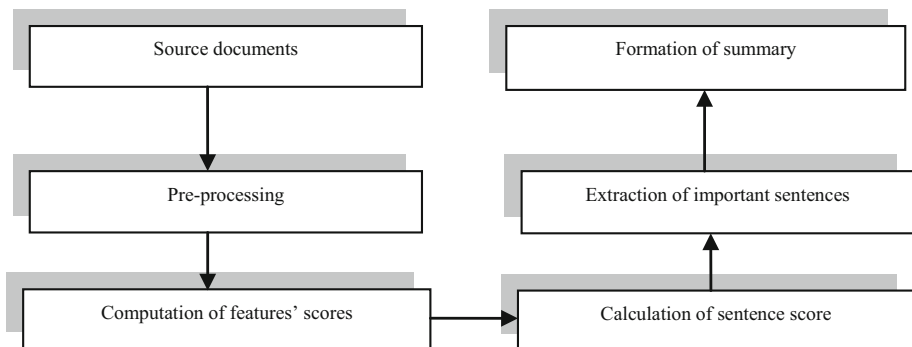
Fig. 1 Block diagram of automatic extractive text summarization system by using statistical techniques

(name entity) in the sentence, node's (sentence's) bushy path, summation of similarities for each node (aggregated similarity), etc. So, for each sentence in the document, a score is computed and highly scored sentences are chosen for generating the summary. Some other features that can discover important words are TF*IDF (Term Frequency–Inverse Document Frequency), information gain, mutual information and residual inverse document frequency. Each of the above features assigns some weight to the words. Based on these weights, the scores are assigned to the sentences and then highly scored sentences are chosen to generate the summary.

Figure 1 above displays the block diagram of automatic text summarization system based on statistical approach. Firstly, preprocessing of the source document is done in which linguistic techniques are applied which includes segmentation of sentences, removal of stop-words, removal of punctuation marks, stemming, etc. Segmentation process deals with dividing the text into sentences. Then, elimination of stop-words is done. Words that occur frequently in

the text but have no contribution in selecting the important sentences, for example prepositions, articles, pronouns, etc are termed as stop-words. They are considered as noisy terms within the text. So, their removal would be very helpful before a natural processing task executes. Then, stemming is performed. Stemming is the process of reducing the words with the same root or stem to a common form, thus removing the variable suffixes (Manning et al. 2008). A few popular and efficient stemming algorithms are of Porter and Lovins. Then some features are selected which will help in extraction of important sentences. These features may be statistical or linguistic or a combination of both. For each sentence, all the selected features' scores are computed and then added together to obtain the score of each sentence. Highly scored sentences are then chosen to form the summary while preserving the original order of sentences in the document. Summary's length depends on the compression rate desired.

3.2 Topic based approaches

Topic is the subject of the document, i.e., what the document is about. In Harabagiu and Lacatusu (2005), structure of topic is defined by topic themes that are represented by events which occur frequently in the collection of documents. In this paper topic is represented in five different ways:

- *Topic signatures*: Lin and Hovy (2000) suggested that a collection of terms are required to express a document's topic.
- *Enhanced topic signatures*: It is same as above except that important relations are discovered between two topic concepts.
- *Thematic signatures*: Documents are first segmented using an algorithm, TextTiling (Hearst 1997). Then, themes are assigned some labels so that they can be ranked later.
- *Modeling the documents' content structure*: it is considered that the texts produced by a content model (for e.g., Hidden Markov Model) describe a given topic.
- *Templates*: specific entities or facts are identified here.

3.3 Graph based approaches

In a graph, text elements (words or sentences) are represented by nodes and edges connect the related text elements (semantically related) together. Erkan and Radev (2004) proposed LexRank which is a summarization system for multiple documents where those selective sentences are shown in the graph which are expected to be a part of the summary. If similarity among two sentences lies above a given limit, then there is a connection between them in the graph. After the network is made, important sentences are selected by the system by carrying out a random walk on the graph. Baralis et al. (2013) proposed GRAPHSUM, a novel and general-purpose summarizer based on graph model which represents correlations among multiple terms by discovering association rules.

3.4 Discourse based approaches

This approach is used in linguistic techniques for automatic text summarization. Discourse relations in the text are discovered here. Discourse relations represent connections between sentences and parts in a text. Mann and Thompson (1988) proposed Rhetorical Structure Theory (RST) in computational linguistics domain to act as a discourse structure. RST has two main aspects: (a) coherent texts contain a few number of units, connected together by rhetorical relations, (b) In coherent texts, there must be some kind of relation between

various parts of the text. Coherence as well as cohesion are the two main challenging issues in text summarization. Linguistic approaches are helpful in understanding the meaning of the document for summary generation.

3.5 Approaches based on machine learning

Machine learning based approaches learn from the data. They can be supervised, unsupervised or semi-supervised. In supervised approach, there is a collection of documents and their respective human-generated summaries such that useful features of sentences can be learnt from them. Supervised or trainable summarizers classify each sentence of the test document either into "summary" or "non-summary" class with the help of a training set of documents. Large amount of labeled or annotated data is needed for learning purpose. Support Vector machine (SVM) (Fattah 2014), Naïve Bayes classification (Fattah 2014), Mathematical Regression (Fattah and Ren 2009), Decision trees, Neural networks (Multilayer Perceptron) (Fattah and Ren 2009), etc are some of the supervised learning algorithms. On the other hand, unsupervised systems do not require any training data. They generate the summary by accessing only the target documents. They try to discover hidden structure in the unlabelled data. Thus, they are suitable for any newly observed data without any advanced modifications. Such systems apply heuristic rules to extract highly relevant sentences and generate a summary. Clustering (Yang et al. 2014), Hidden Markov Model, etc are some of the examples of unsupervised learning techniques. Genetic algorithms (GA) (Mendoza et al. 2014) is also a type of machine learning approach. Genetic algorithm, being a search heuristic works on the process of natural selection. Belonging to the category of evolutionary algorithms, they solve the optimization problems by using approaches based on natural evolution like mutation, inheritance, crossover and selection. Semi-supervised learning techniques require labeled and unlabeled data both to generate an appropriate function or classifier.

4 Recent automatic text summarization extractive approaches

Extractive summarization generates an extract summary by selecting a few relevant sentences from the original document. Summary's length depends on the compression rate. It is a simple and robust method for summarization of text. Here some saliency scores are assigned to sentences in the documents and then highly scored sentences are chosen to generate the summary. This section describes in detail some recent extractive text summarization approaches developed in the last decade.

4.1 Trained summarizer and latent semantic analysis for summarization of text

Yeh et al. (2005) proposed two new techniques for automatic summarization of text: Modified Corpus Based Approach (MCBA) and Latent Semantic Analysis-based TRM technique (LSA+TRM). MCBA, being a trainable summarizer depends on a score function and analyzes important features for generating summaries like Position (Pos), +ve keyword, −ve keyword, Resemblance to the Title (R2T) and Centrality (Cen). For improving this corpus-based approach, two new ideas are utilized: (a) in order to denote the importance of various sentence positions, these sentence positions are ranked, (b) Genetic Algorithm (GA) (Russell and Norvig 1995) trains the score function for obtaining an appropriate combination of feature weights. LSA+TRM approach uses LSA (Landauer et al. 1998; Deerwester et al. 1990) to obtain a document's semantic matrix and builds a relationship map for semantic text by
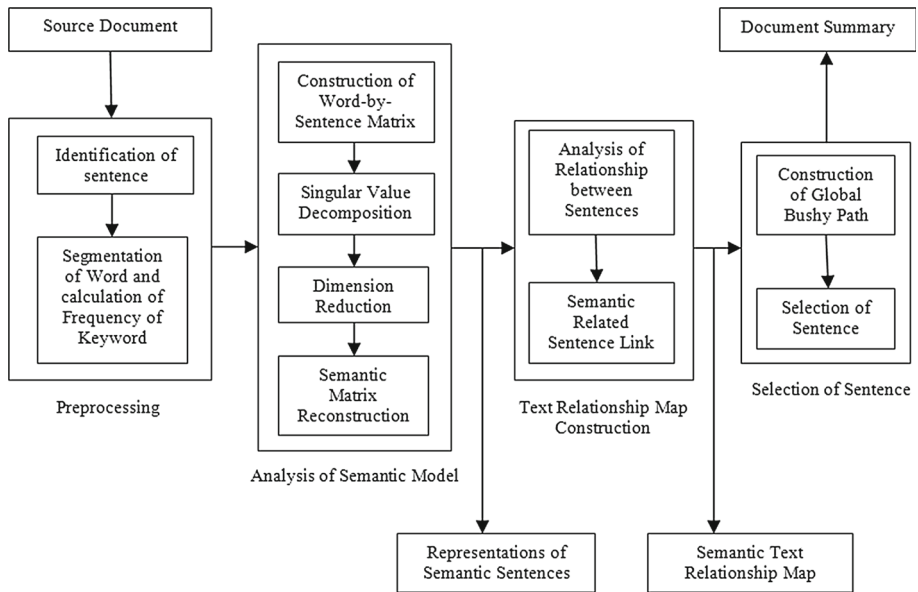
Fig. 2 Complete process of LSA + TRM technique

employing a sentence's semantic representation. LSA is used to extract latent structures from a document. The entire process of LSA + TRM approach shown above in Fig. 2 is divided in four phases. MCBA and LSA + TRM approach focus on summarizing single documents and produce indicative, extract based summaries.

Conclusion: Cen and R2T are the two important features and mix of features like Pos, +ve keyword, Cen and R2T are the best. GA provides an appropriate mix of feature weights during training phase. LSA + TRM performs better than keyword-based text summarization techniques in both single-document as well as corpus level.

4.2 Information extraction using sentence based abstraction technique

Chan (2006) proposed a new quantitative model for creating summary which extracts the sentences from highly relevant portion of the text. Shallow linguistic extraction technique is used in this approach. This approach performs information extraction through sentence based abstraction technique. A discourse network is created for representing discourse that not only includes sentence boundaries but also considers text which is composed of interrelated parts as a single unit instead of isolated sentences in a sequence. In a discourse network, discourse segment is the smallest unit of interaction. In this approach, textual continuity is used for combining the segments together via discourse network. Cohesion and Coherence are the two quantitative coefficients to evaluate the amount of discourse continuity. Connection among sentences in the close segments is represented by cohesion and it is described in a text by pragmatic and semantic relations between sentences and clauses (Quirk et al. 1985). Various factors for cohesion taken into account are: referential cohesion (Kintsch and Van Dijk 1978), lexical cohesion (Halliday and Hasan 1991) and verb cohesion (Haberlandt and Bingham 1978). Coherence is the link among adjacent segments which is not visible in the text. There are two types of coherence: local coherence and global coherence. Rhetorical
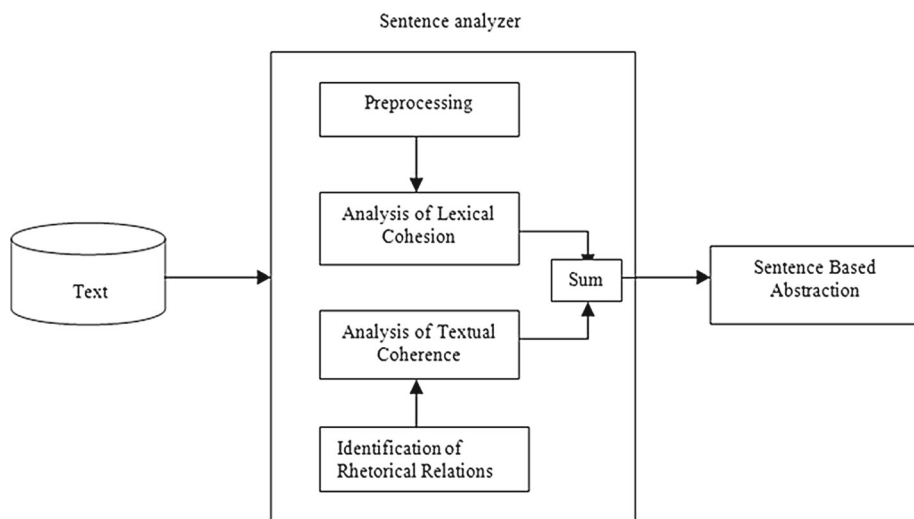
Sentence analyzer

Fig. 3 Schematic diagram of the system

structure Theory (RST) is employed here to model coherence relation in the text. Coherence analysis depends on rhetorical relations (Mann and Thompson 1988). Figure 3 above shows the schematic diagram of the system in which text is first passed through a sentence analyzer and its output is fed to the sentence based abstraction algorithm.

Conclusion: Information retrieval performance is highly improved. Semantically relevant sentences are correlated efficiently.

4.3 Text understanding and summarization through document concept lattice

Ye et al. (2007) proposed a data structure named as Document Concept Lattice (DCL) in which concepts of the source document are represented through a direct acyclic graph such that the set of overlapping concepts are represented by nodes. Here, concepts are words representing concrete entities and their corresponding actions. Thus, concepts indicate important facts and help to answer important questions. Through DCL, the summarization algorithm selects a globally optimum set of sentences that represent maximum number of possible concepts with the use of minimum number of words. This task is accomplished through a fitness metric for a summary that is termed as a summary's representative power. For exploring DCL's search space, dynamic programming is implemented in given three steps: (a) a set of important internal nodes are selected, (b) sentences with highest representative power are selected from these selected important internal nodes and (c) after observing a number of combinations of the chosen sentences, the best combination is selected that leads to the minimum answer loss. Finally, this algorithm produces the output summary with the set of sentences that accounts for the highest representative power.

Conclusion: The proposed approach is competitive with respect to existing sentence-clustering and sentence scoring techniques.

4.4 Sentence extraction through contextual information and statistical based summarization of text

Ko and Seo (2008) proposed an effective method for summarization of text in which important sentences are extracted by applying contextual information and statistical approaches. In this method, initially two consecutive sentences are combined to form a Bi-Gram Pseudo Sentence (BGPS) through sliding window mechanism (Ko and Seo 2004) which solves the feature sparseness problem, caused by obtaining features from a single sentence as BGPS contain a greater number of features (words) than a single sentence. The proposed technique performs sentence extraction tasks of two different types. In the first stage, from the target document, many relevant BGPS are selected. Each selected BGPS is then split into two single sentences. In the second stage, work is done on the separated sentences and in order to produce a final summary, important sentences are extracted. The hybrid statistical sentence extraction methods used here are: title method, location method, aggregation similarity method, frequency method and tf-based query method. The proposed approach is also applied to the multi-document summarization in which there are two sentence extraction processes in which a summary is initially generated for each document in the document cluster via primary process of sentence extraction. Then, from the summaries obtained in the primary process, the resultant summary of the document cluster is produced via secondary process of sentence extraction.

Conclusion: Performance of this method is better than other methods for summarizing both single and multiple documents.

4.5 Summarization of emails through conversational cohesion and subjective opinions

Carenini et al. (2008) proposed new approaches for summarizing email conversations. Initially a fragment quotation graph is built with the conversation involving a few emails in which nodes represent distinct fragments and edges represent replying relationship among fragments. Then this fragment quotation graph helps to form a sentence quotation graph such that a distinct node in this graph represents each sentence in the email conversation and a replying relationship is represented between two nodes by an edge. In order to assign weights to the edges, three kinds of measures for cohesion are explored: clue words (stem-dependent), semantic similarity (WordNet-dependent) and cosine similarity (TF–IDF dependent). The task of extractive summarization is considered as a node ranking problem. Therefore, Generalized ClueWordSummarizer (CWS) (Carenini et al. 2007) and Page-Rank, i.e., the two graph-based summarization approaches, are used for computing each sentence's score (node) and then highly scored sentences are used to generate the summary. In Generalized ClueWord-Summarizer, weight of all outgoing and incoming edges of a node are added to compute the score of a sentence but it does not consider the importance of the node (sentence). Page-Rank based summarizer considers both weights of outgoing and incoming edges along with the importance of nodes (sentences). Subjective opinions are integrated with graph based methods to propose a summarization approach that helps to identify more important sentences. In order to obtain better results, subjective opinions are used with best cohesion measure. The sentence that comprises of more subjective words is considered to be an important sentence for the summary. OpFind (Wilson et al. 2005) and Opbear (Kim and Hovy 2005) are the two lists of subjective words and phrases considered in this approach.

Conclusion: The evaluation results show that the basic CWS (clue word dependent approach) has a better runtime performance and achieves a greater accuracy as compared to the other cohesion metrics. Also, this method has an accuracy greater than that of the Page-Rank algorithm.

4.6 Summarization of text through complex network approach

Antiqueira et al. (2009) proposed a technique based on complex network for extractive summarization of text. This approach employs sentences in a simple network that needs only shallow pre-processing of text. Source text is represented through a network such that each source sentence is represented by a node and an edge is formed by connecting two nodes if their respective sentences have atleast a single common word, i.e., lexical repetition. In the network, there is a limit on the number of edges by considering only lemmatized nouns. In the proposed method, initially pre-processing is performed on the source text in which identification of sentence boundaries is done and lemmatization is performed on nouns. Then, the pre-processed text is arranged in the network representation on the basis of adjacency and weight matrices of order NxN where N is the number of nodes or sentences. The network metrics are evaluated with the help of the above defined matrices and each node is assigned a rank. The n nodes from the beginning of the ranking are chosen to form the summary, where n depends on the compression rate. Seven network measurements (Degree, Shortest path, Locality index, d-rings, k-cores, w-cuts, communities) are used to develop fourteen different summarization strategies which are generically named as CN-Summ. Another summarizer is constructed that works like a voting summarizer, i.e., CN-Voting. This summarizer selects sentences best ranked by the fourteen strategies. The network metrics captures salient text features, thus making text representation through complex networks, a suitable method for automatic summarization.

Conclusion: Results demonstrate that a few CN-Summ versions perform like the Portuguese text summarizers which are reported as best in the literature in terms of informativeness level of the extracts.

4.7 Automatic creation of generic document summaries through non-negative matrix factorization

Lee et al. (2009) suggested a novel unsupervised summarization approach for generic documents through Non-negative Matrix Factorization (NMF). For sentence selection, singular vectors are used in Latent Semantic Analysis (LSA) approach and they could have negative values and they are not sparse so this approach could not capture the meaning of semantic features intuitively which are very sparse and their scope of meaning is narrow. Therefore, LSA based summarization methods are unable to select meaningful sentences (Zha 2002; Lee and Seung 1999). Therefore, in the proposed method, components of semantic feature vectors entirely contain non-negative values and they are also so sparse that semantic features can be interpreted very well. A combination of some relevant semantic features in linear order can be used to represent a sentence. Therefore, subtopics present in a document can be discovered very well and there is an increased chance of extraction of relevant sentences. Using NMF, a method is proposed for selecting sentences to create generic document summaries in which a document is first pre-processed and then summarization is done. NMF is performed on term-by-sentence matrix to produce a non-negative semantic variable matrix. Generic relevance is computed for each sentence which signifies to what extent a sentence discusses

important topics that are represented as semantic features. Then sentences are chosen with highest generic relevance values.

Conclusion: Performance evaluation using *t* test shows that most of the hypothesis is accepted except a few concerning precision. But most F-measure tests are accepted and F-measure is more important than recall and precision. Therefore, NMF shows the best performance among other summarization methods.

4.8 Automatic text summarization using MR, GA, FFNN, GMM and PNN based models

Fattah and Ren (2009) proposed a method to improve selection of content in automatic summarization of text with the help of a few statistical features. This method, being a trainable summarizer focuses on different statistical features in every sentence for producing summaries. These features are: Position of Sentence (Pos), +ve keyword, −ve keyword, Resemblance of sentence to the Title (R2T), Centrality of Sentence (Cen), Presence of Name Entity in sentence (PNE), Presence of Numbers in sentence (PN), Bushy Path of sentence (BP), Relative Length of sentence (RL), and Aggregate Similarity (AS). By combining all these features, Genetic Algorithm (GA) and Mathematical Regression (MR) models are trained for getting an appropriate mix of feature weights. Feed Forward Neural Network (FFNN) and Probabilistic Neural Network (PNN) are used for classification of sentences. Some text features like +ve keyword and −ve keyword are language-dependent while eight others are language-independent. A weighted score function corresponding to a sentence is computed and all the above mentioned features are taken into consideration. All documents' sentences are assigned ranks in decreasing sequence of their scores and a highly scored collection of sentences are employed to produce a summary of the document on the basis of various compression rates (10, 20, 30 % used here).The proposed automatic summarization model has two phases as shown below in Fig. 4.

Conclusion: The results show that feature BP is the most important text feature as it gives best results and feature PND gives the lowest results as numerical data is not present in religious and political articles. GMM approach gave the best results among all techniques as it could model arbitrary densities.
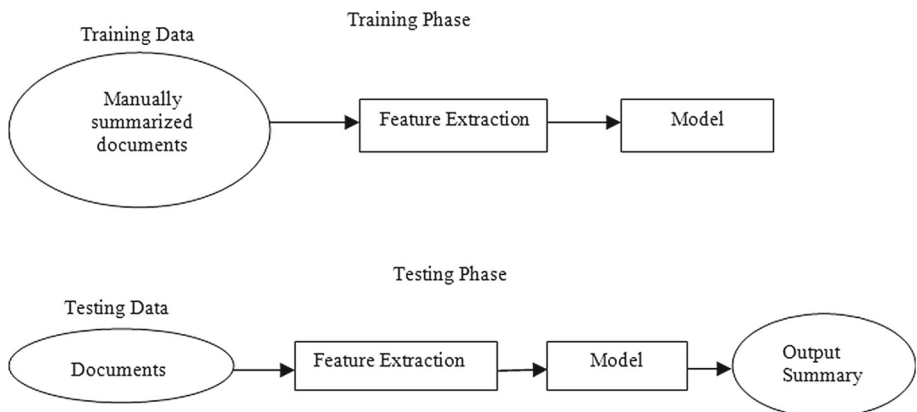

Fig. 4 The proposed automatic summarization model

4.9 Query-based summarization of multiple documents by applying regression models

Ouyang et al. (2011) proposed an approach in which Regression models are applied for ranking the sentences in query-based summarization of multiple documents. In this approach seven features are used to select important sentences in query-based summarization of multiple documents in which three features are query-dependent such as named-entity matching, word-matching and semantic matching and four features are query-independent like sentence position, named entity, word TF–IDF and stop-word penalty. First of all, human summaries create "pseudo" training data. Then, this training data and their set of documents develop and compare various approaches based on N-gram technique which compute "nearly true" relevance scores of sentences. Then, a mapping function is learnt with the help of this training data via collection of features of sentences previously defined. After that relevance of sentences is predicted in the test data through this learned function. For learning regression models, an efficient data set of training data needs these two important things: (a) a suitable set of topics with properly written manual summaries and (b) a suitable method to compute the relevance of sentences. Redundancy is removed from the summary by using Maximal Marginal Relevance (MMR) approach (Carbonell and Goldstein 1998).

Conclusion: Experimental results demonstrate that for computing the importance of sentences, models based on regression outperform learning-to-rank and classification models.

4.10 Maximum coverage and minimum redundancy in summarization of text

Alguliev et al. (2011) proposed an unsupervised summarization model for generic text as an Integer Linear Programming problem (ILP) which directly identifies important sentences from the document as well as consists of the relevant content of the entire document. This approach is named as Maximum Coverage and Minimum Redundancy (MCMR). This approach tries to optimize three important characteristics of a summary: (a) Relevance, (b) Redundancy and (c) Length. A subset of sentences is chosen that covers relevant text of the document collection. Then similarity is computed between the summary and the document collection using NGD-based similarity (Normalized Google Distance) (Cilibrasi and Vitanyi 2007) and cosine similarity and this similarity needs to be maximized. An objective function is defined and needs to be maximized that assures that summary would consist of the important content present in the document collection as well as summary won't have a large number of sentences expressing the same information. At the same time there is a constraint on the length of the summary. Finally an objective function is formed by linearly combining cosine similarity based objective function and NGD-based similarity objective function and this combined objective function also needs to be maximized. This summarization approach is implemented as an optimization problem that tries getting solution to the problem globally. The algorithms that are used to solve ILP problem are: Branch & Bound algorithm (B&B) and Binary Swarm Optimization algorithm.

Conclusion: This approach, i.e., MCMR with B&B algorithm outperforms all other systems. It proves that summarization results rely on the measures of similarity. Through experiments, it is also demonstrated that cosine similarity and NGD-based similarity measures in combination yields results that are better than the results obtained by their separate use.

4.11 Summarization of documents through a progressive technique for selection of sentences

Ouyang et al. (2013) suggested a new progressive method to generate a summary by selecting "novel and salient" sentences. Subsuming relation among two sentences, i.e., asymmetric relation is held among sentences that shows the degree of recommendation of a sentence by some other sentence. In order to determine the relationship between two sentences, relationship between its concepts is discovered that these sentences possess. This relationship between concepts is then accomplished by discovering relationship between words through a coverage-based measure, i.e., a statistical approach that is like the approach used in Sanderson and Croft (1999). All the words occurring in the discovered word relations are organized as a Direct Acyclic Graph (DAG). On the basis of this asymmetric relationship among sentences, progressive technique for sentence selection is developed in which a sentence is either chosen like a novel general sentence or like a supporting sentence. This approach selects new and relevant sentences in the following two ways: (a) uncovered concepts are taken into consideration only during estimation of relevance of the sentences to ensure novelty among sentences and (b) in the meantime relationship between sentences are utilized to enhance saliency measure. In order to execute this technique, from the central node, a random walk is carried out on the DAG to its neighboring nodes such that initially central words are covered and then via word relations, maximum number of words are reached by these covered words. Redundancy is removed by penalizing repetitive words so that each time a new selected sentence brings new concepts.

Conclusion: Progressive system outperforms a typical Sequential system in generation of summaries with better saliency and coverage.

4.12 Evaluation of sentence scoring methods for extractive summarization of text

Ferreira et al. (2013) implemented fifteen scoring methods referenced in the literature in the last decade. Quantitative evaluation is done using ROUGE (Lin 2004) and for qualitative evaluation, the number of sentences are counted that are common between the machine generated summary and the human made summary. Each algorithm's processing time is considered. In order to select relevant sentences, word scoring, sentence scoring and graph scoring methods are used. In word scoring approach, scores are assigned to most important words. Word scoring methods involve word frequency (Luhn 1958; Lloret and Palomar 2009), TF/IDF, upper case (Prasad et al. 2012), proper noun (Fattah and Ren 2009), word co-occurrence (Liu et al. 2009; Tonelli and Pianta 2011; Gupta et al. 2011) and lexical similarity (Murdock 2006; Barrera and Verma 2012; Gupta et al. 2011). In sentence scoring method, features of sentences are analyzed. Sentence scoring methods involve presence of cue-phrases (Kulkarni and Prasad 2010), presence of numerical data (Fattah and Ren 2009), length of sentence, position of sentence (Nobata et al. 2001; Abuobieda et al. 2012; Fattah and Ren 2009) and sentence centrality (Fattah and Ren 2009). In graph scoring approach, scores are computed by observing relationship among sentences. Graph scoring methods include text rank (Barrera and Verma 2012; Mihalcea and Tarau 2004), bushy path of the node and aggregate similarity (Fattah and Ren 2009). Then some suggestions are given to improve sentence scoring results and the six common issues (Orasan 2009) are discussed such as stop words, morphological transformation, similar semantics, ambiguity, redundancy and co-reference.

Conclusion: Results obtained by quantitatively evaluating the summarizers are similar to the results obtained by qualitatively analyzing them through ROUGE. Methods like word frequency and sentence length result in best balance in terms of selection of important sentences and execution time.

4.13 Exploring correlations among multiple terms through a graph-based summarizer, GRAPHSUM

Baralis et al. (2013) proposed GRAPHSUM, a new graph-based, general purpose summarizer for summarization of multiple documents. This approach explores and employs association rules, i.e., a data mining technique for discovering correlations among multiple terms. It does not depend on advanced models based on semantics (like taxonomies or ontologies). After preprocessing, document collection is arranged as a transactional dataset so that association rule mining can be performed on them. Then, frequently occurring itemsets which have high correlations among the terms are extracted from the transactional dataset and a correlation graph is generated from these terms which will further help to select important sentences for the summary. Frequently occurring itemsets are mined by making use of Apriori algorithm and support measure is used for this task. Lift (Tan et al. 2002) measure is used for evaluating positive and negative correlations among the frequently occurring terms and it signifies the strength of association between a pair of terms. The relevance of the graph nodes is estimated by a variant of the traditional PageRank (Brin and Page 1998) graph ranking algorithm. The graph nodes that have a positive correlation with a large number of nodes are placed in the beginning whereas the nodes that have a negative correlation with the surrounding nodes are penalized. Those sentences are chosen for summary generation which best cover the correlation graph as well as have a high relevance score. Greedy algorithm is used for sentence selection over here.

Conclusion: GRAPHSUM outperforms a large number of state-of-art approaches, some of which largely depends on advanced semantic-based models or complex linguistic process.

4.14 Incorporating various levels of language analysis for tackling redundancy in text summarization

Lloret and Palomar (2013) presented an approach to detect redundant information by making use of three methods: lexical, syntactic and semantic levels of language analysis. In lexical based approach, cosine similarity is used to find similarity between sentences of two documents. Those sentences are considered as redundant whose cosine similarity lies above a particular threshold and all those redundant sentences are removed. In syntactic based approach, entailment relations are calculated between pair of sentences to know whether meaning of one sentence can be inferred from the other sentence and the second sentence is taken as redundant if a positive entailment is obtained and as a result it is removed. In semantic based approach, sentence alignment is computed at the document level between a set of related documents through a publically available Champollian Tool Kit.[1] Syntactic and Semantic are better approaches than lexical one where cosine similarity is used. There are two approaches for text summarization. In the first approach, redundant sentences are removed before the text is summarized. The set of non-redundant sentences are then passed to the summarization system in which important sentences are selected by making use of statistical (term frequency) and linguistic (code quantity principle) features and a summary

---

[1] http://champollion.sourceforge.net/.

of predefined length is formed. In the second approach, redundant sentences are employed for summarizing the text because if some information is repeated again and again, then it is considered important and worthy of being a part of the summary. So, a set of redundant sentences obtained from the above given three redundancy detection approaches are input to the summarization system but before inputting them, redundancy is detected using textual entailment. Therefore, all relevant sentences are discovered and at the same time, redundant information is discarded and a summary is generated from the important sentences.

Conclusion: Methods using semantic analysis detect higher redundancy (90 %) whereas redundancy detection decreases with syntactic-based (73 %) or lexical-based approaches (19 %).

4.15 Evolutionary optimization algorithm for summarizing multiple documents

Alguliev et al. (2013) suggested an optimization approach named as OCDsum-SaDE for generic document summarization. This approach deals with content coverage and redundancy at the same time, i.e. it can directly extract important sentences from the given collection, thus covering the relevant portion of the original document and redundancy can be reduced in the summary. An algorithm named as self-adaptive differential evolution (DE) is developed for solving the problem of optimization. One of the key problems while summarizing multiple documents is redundancy. This method focuses on all the three features of summarization: content coverage, diversity and length. Storn and Price (1997) proposed an algorithm based on population named as DE which is similar to Genetic Algorithms (GA) and uses crossover, mutation and selection operators. Search begins in this self-adaptive DE approach with a group of individuals randomly selected from the decision space. Crossover operator is invoked to enhance the diversity of parameter vectors. Mutation operation is used in this algorithm as a method of search and selection operator directs the search towards promising areas in the search space.

Conclusion: The proposed method leads to competitive performance. Statistical results depict that this method performs better than other baseline methods.

4.16 Summarization of multiple documents using a hybrid machine learning model

Fattah (2014) proposed a multi-document summarization approach for enhancing content selection in text by making use of statistical features. A trainable summarizer is used in this approach that employs a number of statistical features like similarity of words among paragraphs (f1), similarity of words among sentences (f2), format of text(f3), score based on the frequency of terms (f4), cue phrases (f5), presence of unimportant information (f6), location of sentence (f7) and title (f8). The text features used here are language-independent. This approach has two modes of operation. Features selected from training data in the training phase help in training Naïve-Bayes Classifier, Maximum Entropy model and Support Vector Machine. Features are computed from sentences in the test data during the testing phase. Weights of features obtained from the training phase help to rank the sentences. In Maximum Entropy approach, a uniform probability distribution is formed and assigned with respect to the feature constraints. This approach is used for classifying the sentences. Naïve-Bayes classifier classifies each sentence as important or unimportant and each sentence is assigned a score. Support Vector Machine helps to obtain an optimal hyper plane separating two classes. Finally a feature vector is used to represent each sentence that can be classified in

any of the two classes, summary or non-summary class. Then the hybrid model formed by combining the above three models is employed to obtain the final sentence ranking.

Conclusion: Obtained results are promising as compared to some existing techniques. Features f1, f2, f5 and f6 give good results.

4.17 Improving clustering at sentence-level with the help of ranking-based technique for theme-based summarization

Yang et al. (2014) proposed a ranking-based sentence clustering framework in which a term is treated like a text object which is independent rather than the feature of a sentence. Clustering of sentences is very important in theme-based summarization where various topic themes are discovered and clusters are based on these themes. Clusters contain highly related sentences. Each theme cluster is based on a generative model, depending on which generative probabilities can be calculated for every target object (object may be a document or term) in each cluster. Figure 5 below shows ranking-based sentence clustering framework.

A probabilistic generative model for sentences is proposed where a set of highly ranked documents and terms are used to generate a sentence. After knowing the generative probabilities for each and every sentence generated from each theme cluster, posterior probabilities are computed for each sentence. Cosine similarity helps in computing similarity between a sentence and a cluster. The above two processes are repeated until sentence clusters do not change remarkably. In the end, allocation of each sentence again occurs to the cluster which is most similar to the sentence. In order to predict the required number of clusters, spectra approach is used that was proposed in Li et al. (2007). After obtaining the sentence clusters, the summaries are produced by selecting the highest ranked sentence from the highest ranked theme cluster to lowest ranked theme cluster, then the second highest ranked sentences from theme clusters in decreasing sequence of their ranks and so on. The method discussed here uses modified MMR-like approach which is an easy and efficient method for selecting summary sentences.
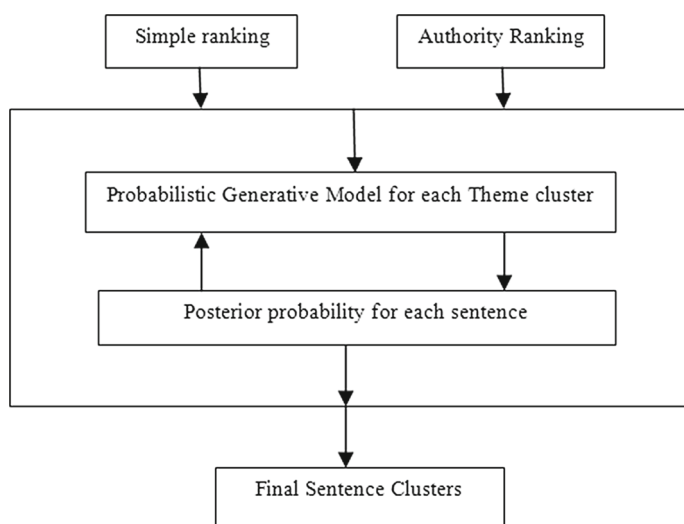

Fig. 5 Ranking-based sentence clustering framework

Conclusion: The proposed framework generates better sentence clusters and gives a better summarization performance.

4.18 Statistical and linguistic based summarization system for multiple documents

Ferreira et al. (2014) suggested a novel graph based clustering algorithm for sentences to tackle with information diversity and redundancy problems in multi-document summarization. The proposed algorithm, along with working on statistical and semantic similarities, linguistically treats input text by performing co-reference resolution and discourse analysis and therefore, develops an unsupervised generic summarization system. This system is based on a four dimensional graph model proposed in Ferreira et al. (2013). Vertices represent the document sentences and there are four distinct relations (semantic similarity, statistical similarity, discourse relations and co-reference resolution) which are represented by four distinct types of edges. Then, TextRank (Mihalcea and Tarau 2004) score for each vertex is computed by making use of the service provided by the summarization model. Vertex that has the highest TextRank score is selected. The user provides a threshold value and by making use of this threshold value and the computed TextRank scores, leader vertices are identified. Each leader vertex represents a cluster. Dijkstra's algorithm (Knuth 1977) is used to compute the path that is shortest between a vertex and each of its leader vertices. The leader vertex closest to each vertex is selected. All those paths that are identified in the previous step in which a vertex is linked to a leader but is distinct from the nearest leader are removed. m graphs are returned by the system representing m clusters such that m is the total number of leader vertices.

Conclusion: When comparing proposed system with its competitors, it achieves 50 % (in the first task with 200 word summary) and 2 % (second task with 400 word summary) better results in terms of F-measure.

4.19 Multi-document summarization based information retrieval using event graphs

Glavaš and Šnajder (2014) presented an event-based summarization and information retrieval model on the basis of event extraction at sentence-level. Event-mentions in text narrate an event's circumstances and are used to represent real-world events. This approach presents three important things: event graphs, event centered information retrieval and summarization models. Vertices represent the event mentions in the text and temporal relationships among the vertices are depicted by edges in an event graph. Supervised machine learning is combined with rule-based models to form a hybrid approach that is used to extract event graphs from the English text. On event graph basis, event-oriented text summarization and information retrieval models for multiple documents are proposed. In event-centered Information Retrieval (IR), documents and query are represented as event graphs. Similarity is measured between a query and a document by holding a comparison using graph kernels among their respective event graphs and documents are then ranked in accordance with the similarity scores, thus obtained. In event-centered summarization model, by considering the relevance of event participants, temporal relations between events and informativeness of events, relevance score is calculated for each event mention in an event graph. Then, scores of event mentions belonging to a sentence are added to obtain scores for each sentence. In order to tackle with redundancy, semantically similar sentences are clustered. Finally, sentences with

highest event based scores are selected from each cluster. These are then sorted in descending order and form a summary of predefined length.

Conclusion: In IR, mixed topic collection performs better than one topic collection in all models. In summarization model, all event graph models perform better than other models.

4.20 Extractive summarization of single documents through genetic operators and guided local search

Mendoza et al. (2014) proposed an extractive generic summarization method for single documents by using generic operators and guided local search. This method uses a memetic algorithm which has combined the population based search of evolutionary algorithm with a guided local search strategy. The task of summarization is treated as a binary optimization problem. Few domain and language independent features are used for searching the important sentences from the documents like position of sentence, resemblance of sentence with the title, sentence length, cohesion and coverage. This algorithm's main objective is to take the search towards the most important regions of the search space by co-operation process that refers to the summarization of new individuals through the exchange of information and population competition that deals with the methods of selecting individuals. In this algorithm, individuals in the population are termed as agents and solutions are represented by these agents. The agents in the population compete and co-operate with each other during evolution. Initially, each agent in the population is assigned a fitness value based on their ability to solve the given problem. For initializing the population, each agent is generated randomly so that each sentence gets a fair chance of being part of the agent (i.e., summary). While generating the agent, the summary length constraint must be satisfied. Here a certain number of agents are selected from the current population using an elitist strategy. Then, one-point crossover, multi-bit mutation and finally optimization are performed on the agent. The optimized agent is included in the population and an agent is selected for replacement according to a specific replacement technique. When a new offspring is generated, convergence of the population is evaluated. If the population converges, the population is re-initialized in the same way as the initialization process of the population while keeping a pre-defined number of best agents from the current population. Memetic algorithm stops executing when the stop condition is met which is the maximum number of evaluations of the objective function. Finally, the summary is generated from the solution vector.

Conclusion: With ROUGE-2, the proposed technique outperforms all other methods on both datasets. But in case of ROUGE-1, the proposed algorithm was outperformed by DE by 6.67 % for dataset DUC 2001 and by UnifiedRank by 0.41 % for dataset DUC 2002. Although in Unified ranking, this method ranks first among all other methods.

4.21 Topic-aspect based summarization through selection of groups

Fang et al. (2015) proposed (TAOS) Topic Aspect-Oriented Summarization which is based on topic factors. These topic factors are various features that describe topics such as capital words are used to represent entity. Various topics can have various aspects and various preferences of features are used to represent various aspects. Based on these topic factors, various groups of features can be extracted and then a group norm penalty and latent variables help to choose common group of features. Latent variable and group norm are included in the task of summarization for the first time. Summary based on one topic describes different aspects. This approach is implemented for text as well as image summarization. Various groups of

features are generated after extracting various types of features from the documents. For text summarization task, word frequency, position and length are the three types of features extracted for each sentence from the document whereas for image summarization color histogram, bag-of-visual word and Histogram of Oriented Gradient (HOG) are used. One feature vector is made by concatenating all these extracted features. The aspects of a topic are described by different groups of features. For word feature, eight feature groups are created like adjective, adverb, verb, noun, pronoun, preposition, wh determiner and symbols and numbers. Similarly for other text features, different feature groups are created. But each image feature is described as an individual feature group as semantic information is not clear for low-level image features. Greedy algorithm is used to generate the summary over here. Two main issues: coverage and diversity are considered here.

Conclusion: TAOS outperforms both text and image summarization approaches.

4.22 Summarization of multiple documents based on social Folksonomy by analyzing semantically

Heu et al. (2015) proposed FoDoSu, a Folksonomy-based system that summarizes multiple documents which exploits Flickr tag clusters for selecting important sentences. Folksonomy is a system of classification generated by creating and managing tags assigned by users. Flickr is a well-known application for tagging pictures that generates tag clusters comprising of a tag and its similar tags. Initially the documents are pre-processed. After preprocessing, the words obtained are used by word analysis module during which a Word Frequency Table (WFT) is constructed. In order to construct a WFT, frequency is computed for each word in the documents and words having high semantic relationships are discovered with the help of tag clusters from Flickr. Words having strong semantic relationship with a particular tag make a tag cluster. The WFT gets updated to WFT' when words having high semantic relationships are discovered. Each word's contribution in the WFT' is calculated by using HITs algorithm with WordCluster. HITS algorithm is used to rate web pages by analyzing links between web pages. WordCluster comprises of a collection of high relationship words in WFT'. After each word's relevance and contribution is analyzed, score of each sentence is computed using rel-gram and each sentence is ranked with WordCluster. The FoDoSu system finally selects highly scored sentences in order to generate summaries of multiple documents.

Conclusion: Results obtained by experiments demonstrate that the efficiency of document summarization is heavily improved by employing Flickr tag clusters as proper nouns and novel words are semantically analyzed by them.

4.23 Other text summarization approaches

A few other most recent text summarization extractive and compressive approaches are explained below. Compressive approaches convert a relevant sentence into a grammatically shorter sentence that preserves the important part of the sentence.

4.23.1 Learning-based approach for summarizing related sentences

Tzouridis et al. (2014) suggested a structured learning-based technique to compress multiple sentences. A word graph is used to represent related sentences such that summaries form the paths in the graph (Filippova 2010). Instead of applying heuristics, dynamic programming has been adapted to the data. Word graphs and compressions are embedded in a joint feature

space where compressions of different quality are learnt to be separated by a generalized linear scoring function. In order to decode the data, a generalized, loss-augmented shortest path algorithm has been developed that is solved through an integer linear program in polynomial time. A large-margin approach is applied for adapting parameterized edge weights to the data such that the shortest path corresponds to the desired summary.

4.23.2 Semantic role labeling with minimal resources

Kaljahi et al. (2014) presented a projection-based approach to Semantic Role Labeling (SRL). A number of experiments are carried out with a small set of manually annotated training data and a large set of French semantic role labels which have been projected from a source language through word alignment. The results show that it is better to train the SRL system with a small set of manually annotated training data because if we increase the number of artificial projections, the performance does not improve as expected. It is also found out that it does not make much difference to French SRL performance whether we use universal part-of-speech tags and syntactic dependencies or the original fine-grained tagset and dependencies. Also, direct translations are no more useful than indirect translations.

4.23.3 Summarizing single documents through nested tree structure

Kikuchi et al. (2014) suggested an approach for summarizing single documents that makes use of dependency between sentences obtained through rhetorical structures and dependency between words obtained through a dependency parser. Both of these dependencies are represented by building a nested tree for a document which is composed of two types of tree structures: a document tree in which nodes represent dependencies between sentences and a sentence tree in which nodes represent dependencies between words. A nested tree is constructed by replacing the nodes in a document tree by a sentence tree. This method extracts a rooted document subtree from a document tree whose nodes are arbitrary subtrees of the sentence tree. The summarization task is formulated as an integer linear programming problem that trims the nested tree without losing important content in the source document.

4.23.4 Two-level sparse representation model for summarization of multiple documents

Liu et al. (2015) proposed MDS-Sparse, a two-level sparse representation model for multi document summarization that employs document reconstruction and is based on three important properties of an ideal reconstructable summary: coverage, sparsity and diversity. At level-1, the set of all summary sentences is sparsely represented by the original document set and at level-2, all sentences present in the original document set are sparsely reconstructed by the summary set. This model is NP-hard and uses simulated annealing algorithm to achieve summarization. Each sentence in the original document set is represented as a non-negative linear combination of only some summary sentences.

4.23.5 Sparse-coding based reader-aware summarization system for multiple documents

Li et al. (2015a, b) proposed reader-aware summarization system for multiple documents (RA-MDS) based on sparse-coding technique that generates summaries not only from the reports of the events but also considers the reader comments at the same time. This system

also aims at improving the linguistic quality of the summary through entity-rewriting. The proposed system is a compression-based unified optimization framework that generates compressive summaries by working at a finer syntactic level known as noun or verb phrase. A dataset is also generated for the summarization task.

4.23.6 Summarization of multiple documents through recursive neural networks based ranking approach

Cao et al. (2015a) proposed Recursive Neural Networks (RNN) based ranking approach for ranking sentences in order to summarize multiple documents. Ranking of sentences is done through a hierarchical regression process which evaluates the relevance of a sentence (non-terminal node) in the parsing tree. On the basis of supervisions from word-level to sentence-level, recursive neural networks are automatically used to learn ranking features over the tree with inputs as hand-crafted feature vectors of words. Ranking scores of words and sentences are used to select important and non-redundant sentences to form summaries. Two methods are used here for sentence selection: greedy algorithm and integer linear programming (ILP).

4.23.7 Graph-based extractive summarization by considering importance, non-redundancy and coherence

Parveen and Strube (2015) proposed an extractive, graph-based unsupervised technique for summarizing single documents which considers three important properties of summarization, i.e. importance, non-redundancy and local coherence. Input document is represented by means of a bipartite graph consisting of sentences and entity nodes. A graph based ranking approach is implemented on this graph for computing rank of the sentences based on their importance. The summary is made non-redundant and locally coherent through the process of optimization.

4.23.8 Sparse optimization based compressive document summarization

Yao et al. (2015a, b) proposed sparse optimization based extractive document summarization which has a decomposable convex objective function that is solved by an efficient alternating direction method of multipliers (ADMM) algorithm. In order to achieve diversity in the summary sentences, an additional sentence dissimilarity term is introduced in the optimization framework. Then the proposed framework is generalized to compressive summarization and a block co-ordinate descent algorithm is derived along with recursive dependency tree compression to optimize the objective function.

4.23.9 Submodular mixtures based summarization of multi-document hierarchy of topics

Bairi et al. (2015) suggested an approach that depends on a family of submodular functions for summarizing topics for a set of documents through DAG-structured hierarchy of topics. Suitable topics are selected by considering properties like coverage, diversity, specificity, clarity and relevance. This approach is based on submodular maximization and structured prediction methods are explored for learning weighted mixtures of submodular functions. The proposed technique can directly incorporate outputs from other algorithms such as LDA, classification and clustering. For evaluation purpose, Wikipedia disambiguation pages are

generated automatically for a set of articles by employing human generated clusterings as ground truth.

4.23.10 Disaster summarization through prediction of salient updates

Kedzie et al. (2015) proposed an update summarization approach that monitors events across time. This approach predicts the importance of sentences with respect to a disastrous event through some disaster specific features such as language model, geographic relevance and temporal relevance based features and also through a few basic and query based features. Then, these predictions are combined with clustering based summarization system for multiple documents. Finally, the most novel and relevant sentences describing the event are selected, thus enhancing the quality of updates.

4.23.11 Summarizing multiple documents through system combination

Hong et al. (2015) suggested an approach for summarizing multiple documents in which summaries generated from different systems are combined. In this approach, initially four portable unsupervised systems are employed to generate basic summaries. Then these basic summaries are combined on sentence level to generate candidate summaries. Finally a supervised model is employed to select among the candidate summaries via utilizing a rich collection of features that can capture important content from different perspectives.

4.23.12 Phrase-based compressive cross-language summarization

Yao et al. (2015a, b) proposed a phrase-based cross-language document summarization system that is able to translate the source documents into a summary in a different language. The scoring function employed in this approach is based on phrase-based machine translation models. Sentence scoring, extraction and compression are performed simultaneously through a phrase-based model, designed in this approach. An efficient greedy algorithm is used to approximately optimize the scoring function. This system translates DUC 2001 English documents to Chinese summaries.

4.23.13 Re-evaluation of automatic summarization using BLEU and 192 variants of ROUGE

Graham (2015) analyzed evaluation of summarization systems by using a machine translation metric, BLEU and 192 variants of ROUGE. Performance of various variants of metrics is evaluated by finding correlations with human assessment. Williams test is used here to test significance of difference between the performance of competing summarization metrics. Results disclose that superior variants of metrics are different from the ones best recommended previously. Recent evaluation of state-of-the-art summarization systems is replicated which also reveals distinct conclusions about the relative performance of systems, showing that precision-based BLEU is on par with recall-based ROUGE.

All the above explained recent text summarization approaches are listed in Table 2 below with their specific need and pros and cons, respectively.

Table 2 Need and pros and cons of recent automatic text summarization extractive approaches

Summarization approach	Need of the approach and their pros and cons
Trained summarizer and latent semantic analysis for summarization of text (Yeh et al. 2005)	*Need*: MCBA + GA can be used to work with the corpus of a particular domain and also for on-line purpose. On the other hand, LSA + TRM is suitable when quality of summary is the top priority
	Pros: LSA + TRM approach generates a summary of semantically related sentences. The approaches are language-independent
	Cons: Summaries lack coherence and cohesion most of the times. Feature weights of score function generated by GA doesn't always give good performance results for the test corpus. In LSA + TRM approach, obtaining the best dimension reduction ratio and explaining LSA effects are difficult. Moreover, it takes a large time to compute SVD
Information extraction using sentence based abstraction technique (Chan 2006)	*Need*: This approach is used to generate summary with semantically related sentences by focusing on factors of textual continuity like textual coherence and lexical cohesion
	Pros: Documents at a higher level can be easily understood by this approach and it depicts human perception better. It represents text better than keywords or cue-phrases and it is also helpful in improving retrieval performance
	Cons: In this approach, only causal coherence is considered whereas causality, temporality and spatiality which are inter-related links for promoting global coherence, are also required for representing behavioral episodes in a discourse
Text understanding and summarization through document concept lattice (Ye et al. 2007)	*Need*: This approach is required to generate a summary of coherent sentences with minimal answer loss, to focus more on semantics and to cover all distinct and relevant local topics via concepts with minimum number of words
	Pros: In this approach topics are represented in a simple way as open-class words and phrases. DCL focuses on semantics and employs only reliable features. The sentences are coherent and represent important and different local topics and the summary is generated with least loss of answer. The evaluation framework does not require human-made summaries
	Cons: Computation cost is more for generating a complete DCL because of observing all possible combinations of concepts
Sentence extraction through contextual information and statistical based summarization of text (Ko and Seo 2008)	*Need*: This approach is required when documents of different languages need to be summarized as this approach is language-independent
	Pros: The biggest strength of this method is that it is language-independent. It solves feature sparseness problem. It doesn't require much processor and memory capacity for extracting important sentences. Also documents without title can be summarized with this approach
Summarization of emails through conversational cohesion and subjective opinions (Carenini et al. 2008)	*Need:* This approach is required to summarize emails taking into consideration conversational structure among emails and the subjective words and phrases they contain
	Pros: Emails can be summarized which helps users have a quick view through the previous discussions via emails in a very short period of time. By integrating subjective opinions into the system, efficiency of the system is further improved

Table 2 continued

Summarization approach	Need of the approach and their pros and cons
Summarization of text through complex network approach (Antiqueira et al. 2009)	*Need*: This approach can be used when it is required to summarize documents of different languages and a sufficient number of linguistic resources are not available
	Pros: This approach is language-independent. Extracts are generated by only using shallow linguistic knowledge. These complex network concepts provide different complementary views of a network
Automatic creation of generic document summaries through non-negative matrix factorization (Lee et al. 2009)	*Need:* This approach is required to perform generic text summarization when training summaries are not available to train the system and also when semantic features need to be explored efficiently
	Pros: This approach extracts more meaningful sentences and subtopics present in the document can be discovered efficiently. It doesn't require any training data as it is an unsupervised approach
Automatic text summarization using MR, GA, FFNN, GMM and PNN based models (Fattah and Ren 2009)	*Need*: This approach is required if we need to use statistical methods for text summarization. Also when we need to employ a trainable summarizer and we have training summaries in a particular language but we want to summarize documents in another language
	Pros: By this approach, the models can be initially trained on data of certain language and can then be tested on data of another language. All the features employed here are language independent except positive keyword and negative keyword
Query-based summarization of multiple documents by applying regression models (Ouyang et al. 2011)	*Need*: This approach is required when it is needed to perform machine learning based summarization of multiple documents by employing query-based features. It is also required when it is needed to develop pseudo training data sets from human summaries to estimate the score of sentences
	Pros: Regression models produce better results than classification and learning to rank methods. A better mapping function is generated between feature vectors and their sentence relevance scores. Redundancy is also removed from summaries by using MMR approach
Maximum coverage and minimum redundancy in summarization of text (Alguliev et al. 2011)	*Need*: This approach can be implemented when there is a need to use an optimization approach for solving the problem of summarization. Also when there are not enough training summaries and when the aim is to cover important content of the document with minimum redundancy
	Pros: This approach is an unsupervised generic text summarization approach so it does not require training summaries. It can generate a summary consisting of relevant content with minimum redundancy
Summarization of documents through a progressive technique for selection of sentences (Ouyang et al. 2013)	*Need*: This approach needs to be implemented in multi document summarization when saliency, novelty and coverage of concepts is a more important issue and when we need to deal with subsuming relationship between sentences by identifying word relations
	Pros: This approach can detect a large number of important concepts than previous techniques. Also novelty of concepts is assured by controlling redundancy

Table 2 continued

Summarization approach	Need of the approach and their pros and cons
Evaluation of sentence scoring methods for extractive summarization of text (Ferreira et al. 2013)	*Need*: This approach needs to be referred to have an insight into the literature in the last decade for getting familiar with the various strategies of text summarization and to know how qualitative and quantitative assessment is performed on fifteen algorithms of sentence scoring
	Pros: Through this approach we get familiar with qualitative and quantitative assessment of sentence scoring algorithms and also some useful directions are provided for improving sentence scoring results
Exploring correlations among multiple terms through a graph-based summarizer, GRAPHSUM (Baralis et al. 2013)	*Need*: This approach is required to discover correlations between several terms present in the document by employing association rules
	Pros: This technique does not depend on advanced semantic-based models and perform a minimum number of language-dependent tasks so it is flexible and portable and can be used with documents belonging to different application contexts
Incorporating various levels of language analysis for tackling redundancy in text summarization (Lloret and Palomar 2013)	*Need*: This approach is required to detect redundancy through three distinct levels of language analysis like syntactic, lexical and semantic
	Pros: This approach discards redundant information and generates a summary from non-redundant information and also this redundant information helps in detecting important sentences
Evolutionary optimization algorithm for summarizing multiple documents (Alguliev et al. 2013)	*Need*: This approach is required to perform optimization-based generic document summarization and to achieve maximum coverage of content with minimum redundant information
	Pros: This approach reduces redundancy in the summaries, selects important sentences from the document and includes relevant content of the original document
	Cons: Runtime complexity of DE which is a population-based stochastic search method is more
Summarization of multiple documents using a hybrid machine learning model (Fattah 2014)	*Need*: This approach is needed to generate summaries of different languages and it can also be used when there is availability of training summaries to implement a trainable summarizer developed from a combination of machine learning algorithms
	Pros: All the text features used in this approach are language-independent. The feature extraction criteria used in this approach provides an opportunity to employ a number of variations on the basis of language and text type
Improving clustering at sentence-level with the help of ranking-based technique for theme-based summarization (Yang et al. 2014)	*Need*: This approach is required to perform theme based multi-document summarization such that sentences are clustered on the basis of themes
	Pros: This approach generates high quality sentence clusters based on theme and a modified MMR-like approach is used to control redundancy in multi-document summarization

Table 2 continued

Summarization approach	Need of the approach and their pros and cons
Statistical and linguistic based summarization system for multiple documents (Ferreira et al. 2014)	*Need*: This approach is required to perform multi-document generic text summarization where there is insufficient availability of well-formed training summaries and when it is required to use statistical, semantic and linguistic, all types of information on the input text
	Pros: Apart from using statistical and semantic similarities, this approach linguistically treats the input text by performing discourse analysis and co-reference resolution. As it is an unsupervised approach, it does not require annotated corpus
	Cons: This system strives to search for important sentences in groups of different topics and hence suffers from the problem of sentence ordering
Multi-document summarization based information retrieval using event graphs (Glavaš and Šnajder 2014)	*Need*: This summarization approach is required to work on domains consisting of real world events like police reports, news stories, etc
	Pros: Event graphs used in this approach cover all important information about real world events and they not only contain temporal information but also semantic information about the events
	Cons: The models which have been proposed are not suitable for descriptive text like art reviews that consists of a very few number of event mentions
Extractive summarization of single documents through genetic operators and guided local search (Mendoza et al. 2014)	*Need*: This approach is required for single-document language-independent generic text summarization. This technique needs to be implemented for directing the exploration towards the most promising regions of the search space
	Pros: All features are domain and language-independent. Memetic algorithm exploits the problem knowledge and redirects the search towards a best solution. Multi-bit mutation encourages information diversity
Topic-aspect based summarization through selection of groups (Fang et al. 2015)	*Need*: This approach is required to discover various aspect preferences and create summaries accordingly for distinct topics
	Pros: Addition of group selection enhances performance of summarization. Coverage and diversity both are considered here
Summarization of multiple documents based on social Folksonomy by analyzing semantically (Heu et al. 2015)	*Need:* The proposed approach is required because approaches using WordNet are unable to analyze proper nouns and novel words as WordNet doesn't cover such words. Therefore, Flickr tag clusters are explored to semantically analyze novel words and proper nouns present in the documents
	Pros: The proposed system has a low computational cost for semantically analyzing words in the document. Also, this method employs Flickr tag clusters that semantically analyze novel words and proper nouns present in the documents which approaches using WordNet fail to analyze

Table 2 continued

Summarization approach	Need of the approach and their pros and cons
Learning-based approach for summarizing related sentences (Tzouridis et al. 2014)	*Need:* This multi-sentence compression approach is required to simplify the summaries as it maps a set of related sentences to a grammatical short sentence that retains the most important information
	Pros: Only a set of five features are sufficient for improving the performance of this approach as compared to the other graph-based multi-sentence compression techniques found in the literature
Semantic Role labeling with minimal resources (Kaljahi et al. 2014)	*Need:* This approach is required to know that a small set of manually annotated training data performs better than a large set of French semantic role labels, projected from a source language
	Pros: This approach suggests that there is no need to generate a large amount of artificial data to train an SRL system
Summarizing single documents through nested tree structure (Kikuchi et al. 2014)	*Need:* This approach is required to generate coherent summaries. This approach is also required when it is needed to jointly utilize relations between sentences and relations between words
	Pros: This approach considers both dependency between words and dependency between sentences at the same time by developing a nested tree. The summaries that are generated are coherent
Two-level sparse representation model for summarization of multiple documents (Liu et al. 2015)	*Need:* This approach is required to generate a reconstructable summary of multiple documents by conserving three important properties: coverage, sparsity and diversity
	Pros: This approach considers document reconstruction problem that by default contains diversity. The speed of this method is also competitive
Sparse-coding based reader-aware summarization system for multiple documents (Li et al. 2015a, b)	*Need:* This approach is required to generate summaries by considering both news reports as well as reader comments for the events
	Pros: This approach uses sparse-coding technique that selects sparse and diverse semantic units. Also the summary, thus generated has a higher linguistic quality by employing the process of entity rewriting
Summarization of multiple documents through recursive neural networks based ranking approach (Cao et al. 2015a)	*Need:* This approach is required to employ the learning ability of recursive neural networks as this approach can automatically learn ranking features over the tree
	Pros: This approach can effectively learn ranking features of sentences and words over the parsing tree, thus providing efficient ranking scores to words and sentences. Sentence selection method used in this approach is more accurate
Graph-based Extractive summarization by considering importance, non-redundancy and coherence (Parveen and Strube 2015)	*Need:* This approach is required to generate non-redundant and locally coherent summaries from documents of different domains and genres
	Pros: This approach doesn't depend on any parameter and training data as it is an unsupervised technique and summary, being coherent is of good quality
	Cons: This approach is capable of generating summary from a single document only

Table 2 continued

Summarization approach	Need of the approach and their pros and cons
Sparse optimization based compressive document summarization (Yao et al. 2015a, b)	*Need:* This approach is required to achieve compressive summarization that yields better results as compared to original extractive systems based on data reconstruction
	Pros: The proposed method is entirely unsupervised so it requires no training data
Submodular Mixtures based summarization of multi-document hierarchy of topics (Bairi et al. 2015)	*Need:* This approach is required to generate Wikipedia disambiguation pages for a set of articles based on different topics but with similar titles
	Pros: This approach can summarize large collection of labels into smaller, manageable and more meaningful sets of labels
Disaster Summarization through prediction of salient updates (Kedzie et al. 2015)	*Need:* This approach is especially required to generate updates across time describing disastrous events which can serve the information needs of responders, crisis management organizations and victims
	Pros: The proposed approach that combines salience with clustering generates more relevant summaries than the approaches employing clustering or salience separately thus helping to share appropriate information in time during a disastrous event
Summarizing multiple documents through system combination (Hong et al. 2015)	*Need:* The proposed approach is required as it can greatly improve the content quality by combining summaries generated from different systems
	Pros: The proposed approach of combining summaries from different systems helps to improve the content quality. Also, this approach can combine summaries generated by any systems
Phrase-based compressive cross-language summarization (Yao et al. 2015a, b)	*Need:* This approach is required to help readers get the main idea of the documents written in a particular language that they are not familiar with
	Pros: Despite of not using any syntactic information, the proposed system maintains better grammaticality and fluency
Re-evaluation of automatic summarization using BLEU and 192 variants of ROUGE (Graham 2015)	*Need:* The evaluation of summarization system is carried out to know which variant of summarization metric significantly outperforms others
	Pros: The evaluation results rectified the wrong assumptions of readers as the results show that superior variants of summarization metrics are different from the ones best recommended previously

5 Comparison of recent automatic text summarization extractive approaches

The techniques that have been explained above in Sect. 4 are compared in a tabular form with some more details about them. Table 3 below shows such a comparison of these extractive text summarization techniques.

Table 3 Comparison of recent automatic text summarization extractive approaches

Summarization approach and its author	Dataset used	Evaluation measure	Baseline approaches	Results
Trained summarizer and latent semantic analysis for text summarization (Yeh et al. 2005)	100 political articles from New Taiwan Weekly[a]	Precision, Recall and F-measure	CBA and MCBA, MCBA and MCBA+GA, LSA+TRM and keyword+TRM	For MCBA+GA, Recall=Precision=F-measure=0.5151 and for LSA+TRM, Recall=Precision=F-measure=0.4442 in single document level
Information extraction using sentence based abstraction technique (Chan 2006)	Four Texts of Stein and Glenn	Spearsman Rank Correlation Coefficient	Casual Chains formed in Trabasso's Experiment	For Text 1=0.42, Text 2=0.51, Text 3=0.65 and Text 4=0.45
Text understanding and summarization through document concept lattice (Ye et al. 2007)	DUC 2005 and DUC 2006	ROUGE-2 and ROUGE-SU4	Techniques available for clustering and scoring of sentences	For DUC 2005, R-2 Recall=7.17% and R-SU4 Recall=13.16% and in DUC 2006, second best ROUGE scores are obtained, i.e. R-2 Recall=8.99% and R-SU4 Recall=14.75%
Sentence extraction through contextual information and statistical based summarization of text (Ko and Seo 2008)	KOrea Research and Development Information Center (KORDIC) dataset and news articles	Precision, Recall and F_1-measure	Title, Location, MS Word and DOCUSUM (Ko et al. 2003)	F_1 score=55.3 (with title) for single document summarization and F_1 score=51.6 for multi document summarization
Summarization of emails through conversational cohesion and subjective opinions (Carenini et al. 2008)	Enron email dataset	Sentence pyramid precision, ROUGE-2 and ROUGE-L	CWS, CWS-Cosine, CWS-lesk, CWS-jcn, PR-Clue, PR-lesk, PR-Cosine, PR-jcn, OpFind, OpBear	For CWS+OpFind, pyramid precision=0.65, R-2=0.50, R-L=0.60. For CSW+OpBear, pyramid precision=0.64, R-2=0.49, R-L=0.59

Table 3 continued

Summarization approach and its author	Dataset used	Evaluation measure	Baseline approaches	Results
Summarization of text through complex network approach (Antiqueira et al. 2009)	TeMario corpus that consisted of 100 news articles in Brazilian Portuguese	Precision, Recall, F-measure and ROUGE-1	Top Baseline and Random Baseline and six other extractive systems i.e., ClassSumm (Neto et al. 2002), NeuralSumm (Pardo et al. 2003a), GistSumm (Pardo et al. 2003b), TF-ISF-Summ (Neto et al. 2000), SuPor (Rino and Modolo 2004) and its improved version SuPor-v2 (Leite and Rino 2006)	Precision = 48.1, Recall = 40.3, F-measure = 42.9 and ROUGE-1 = 0.5031
Automatic creation of generic document summaries through non-negative matrix factorization (Lee et al. 2009)	DUC 2006	ROUGE	RM (Gong and Liu 2001), LSA, MRP (Zha 2002) and LGP (Kruengkrai and Jaruskulchai 2003)	Recall values for R-1 = 0.2763, R-L = 0.2541, R-W = 0.0732 and R-SU = 0.0853
Automatic text summarization using MR, GA, FFNN, GMM and PNN based models (Fattah and Ren 2009)	200 Arabic articles related to politics and 150 English articles related to religion and DUC 2001	Precision	GA, MR, FFNN, PNN and GMM	For DUC 2001, Precision scores for the following methods are: GA = 0.4335, MR = 0.4021, FFNN = 0.4423, PNN = 0.4543 and GMM = 0.6046
Query-based summarization of multiple documents by applying regression models (Ouyang et al. 2011)	DUC 2005, DUC 2006 and DUC 2007	ROUGE-2 and ROUGE-SU4	Learning-to-rank and Classification models. Also with human summarizers and DUC systems that perform best	For Uni+Max, DUC 2005 evaluation results are: R-2 = 0.0757 R-SU4 = 0.1335. For DUC 2006, R-2 = 0.0926, R-SU4 = 0.1485. For DUC 2007, R-2 = 0.1133, R-SU4 = 0.1652

Table 3 continued

Summarization approach and its author	Dataset used	Evaluation measure	Baseline approaches	Results
Maximum coverage and minimum redundancy in summarization of text (Alguliev et al. 2011)	DUC 2005 and DUC 2007	ROUGE-2 and ROUGE-SU4	For DUC 2005, six methods were employed, i.e., TranSumm (Zhao et al. 2009; Amini and Usunier 2009), Biased LexRank (Otterbacher et al. 2009), Content-term(He et al. 2008), Qs-MRC (Wei et al. 2008) and TMR+TF (Tang et al. 2009). For DUC 2007, four methods were used, i.e., PNR² (Wenjie et al. 2008), GSPSum (Zhang et al. 2008a), PPRSum (Liu et al. 2008) and AdaSum (Zhang et al. 2008b)	On DUC 2005, for MCMR (B&B), R-2=0.0790, R-SU4=0.1392 and for MCMR (PSO), R-2=0.0754, R-SU4=0.1360. On DUC 2007, for MCMR (B&B), R-2=0.1221, R-SU4=0.1753 and for MCMR (PSO), R-2=0.1165, R-SU4=0.1697
Summarization of documents through a progressive technique for selection of sentences (Ouyang et al. 2013)	DUC 2004, DUC 2005, DUC 2006 and DUC 2007	ROUGE-1 and ROUGE-2	Typical Sequential summarization System	On DUC 2004, for progressive approach, R-1=0.519, R-2=0.147. For progressive query-based approach, on DUC 2005, R-1=0.393, R-2=0.080. On DUC 2006, R-1=0.419, R-2=0.095. On DUC 2007, R-1=0.443, R-2=0.122
Evaluation of sentence scoring methods for extractive summarization of text (Ferreira et al. 2013)	CNN Dataset for news articles, Blog summarization dataset for blogs and SUMMAC dataset for articles context	For Quantitative assessment, ROUGE was used and Qualitative assessment was carried out by four people	Fifteen methods of sentence scoring present in the literature	Results obtained by quantitatively evaluating the summarizers are similar to the results obtained by qualitatively analyzing them through ROUGE. Methods like word frequency and sentence length result in best balance in terms of selection of important sentences and execution time

Table 3 continued

Summarization approach and its author	Dataset used	Evaluation measure	Baseline approaches	Results
Exploring correlations among multiple terms through a graph-based summarizer, GRAPHSUM (Baralis et al. 2013)	DUC 2004 and five real-life collections of news document	ROUGE-2 and ROUGE-SU4	Thirty-five summarizers submitted to DUC 2004, eight summaries made by humans, two open source text summarizers: OTS (Rotem 2011) and Texlexan (2011)[b] and ItemSum (Baralis et al. 2012)	On DUC 2004, for R-2, Recall=0.093, Precision=0.099, F-measure=0.097 and for R-SU4, Recall=0.015, Precision=0.021, F-measure=0.019
Incorporating various levels of language analysis for tackling redundancy in text summarization (Lloret and Palomar 2013)	DUC 2002, DUC 2003 and DUC 2004	ROUGE-1, ROUGE-2 and ROUGE-L	LEADBASED, RANDOM, MEAD-CoSim, MEAD-MMR	Methods using semantic analysis detect higher redundancy (90%) whereas redundancy detection decreases with syntactic-based (73%) or lexical-based approaches (19%)
Evolutionary optimization algorithm for summarizing multiple documents (Alguliev et al. 2013)	DUC 2002 and DUC 2004	ROUGE-1, ROUGE-2, ROUGE-L and ROUGE-SU	DUCbest, Random, FGB (Wang et al. 2011), NMF (Lee et al. 2009), LSA (Gong and Liu 2001), BSTM (Wang et al. 2009), LexRank (Erkan and Radev 2004), Centroid (Radev et al. 2004a, b), MCKP (Takamura and Okumura 2009), WFS-NMF (Wang et al. 2010) and WCS (Wang and Li 2012)	On DUC 2002, R-1=0.4990, R-2=0.2548, R-L=0.4708, R-SU=0.2855. On DUC 2004, R-1=0.3954, R-2=0.0969, R-L=0.3927, R-SU=0.1367
Summarization of multiple documents using a hybrid machine learning model (Fattah 2014)	DUC 2002	ROUGE-1	Lead Baseline approach, UnifiedRank, PositionRank, TwoStageRank and CLASSY's guided summarization	R-1=0.3862

Table 3 continued

Summarization approach and its author	Dataset used	Evaluation measure	Baseline approaches	Results
Improving clustering at sentence-level with the help of ranking-based technique for theme-based summarization (Yang et al. 2014)	DUC 2004 and DUC 2007	Cluster quality in terms of Modularity measure and ROUGE-1, ROUGE-2 and ROUGE–SU4	Interactive, Integrated, Context-based, LSA-based, WordNet-based and Word-based summarization model	Cluster quality for Ranking-based system on DUC 2004 is 0.579 and on DUC 2007 is 0.661. For Ranking-based MMR system, on DUC 2004, R-1=0.37878, R-2=0.09357, R-SU4=0.13253 and on DUC 2007, R-1=0.44221, R-2=0.12618, R-SU4=0.17802
Statistical and linguistic based summarization system for multiple documents (Ferreira et al. 2014)	DUC 2002	F-measure	DUC 2002 systems (System 24, System 19, System 20, System 29, System 28) with 200 word summary	On DUC 2004, for the first task (200 word summary), F-measure=30% and for the second task (400 word summary), F-measure=25.4%
Multi-document summarization based information retrieval using event graphs (Glavaš and Šnajder 2014)	For IR, two test collections were developed each containing fifty queries: mixed topic collection and Topic-specific collection. For summarization, DUC 2002 and DUC 2004	Mean Average Precision (MAP) for IR. For summarization, ROUGE-1 and ROUGE-2	For IR, TF–IDF Vector Space Model, Hiemstra Language Model and two probabilistic models: DRF_BM25 and In_expC2 (Amati 2003). For summarization, best and median performing models from respective shared tasks and human performance	For IR MAP=0.502 for mixed topic collection and MAP=0.407 for one topic collection. For summarization on DUC 2002, R-1=0.415, R-2=0.116 and on DUC 2004, R-1=0.405, R-2=0.107
Extractive summarization of single documents through genetic operators and guided local search (Mendoza et al. 2014)	DUC 2001 and DUC 2002	ROUGE-1 and ROUGE-2	UnifiedRank (Wan 2010), DE (Aliguliyev 2009), FEOM, NetSum (Svore et al. 2007), CRF (Shen et al. 2007), QCS (Wan 2008), SVM (Yeh et al. 2005) and Manifold Ranking (Wan 2010)	On DUC 2001, R-1=0.44862, R-2=0.20142 and on DUC 2002, R-1=0.48280, R-2=0.22840

Table 3 continued

Summarization approach and its author	Dataset used	Evaluation measure	Baseline approaches	Results
Topic-aspect based summarization through selection of groups (Fang et al. 2015)	For text summarization, DUC 2003 and DUC 2004. For image summarization, NUS-Wide dataset	ROUGE-1 and ROUGE-L for text summarization and Jensen-Shannon Divergence (Lin 2004) for image summarization	For text summarization, DSDR (He et al. 2012) and Bud-sub (Lin and Bilmes 2010) (unsupervised methods) and Sub-SVM (supervised approach). For image summarization, unsupervised methods like AP (Frey and Dueck 2007), k-medoids (Hadi et al. 2006) and DL (Yang et al. 2013)	On DUC 2003 with stop-words, R-1=0.40146, R-L=0.35830 and without stop-words, R-1=0.31990. R-L=0.29389. On DUC 2004 with stop-words, R-1=0.41849, R-L=0.36678 and without stop-words, R-1=0.33743, R-L=0.30706
Summarization of multiple documents based on social Folksonomy by analyzing semantically (Heu et al. 2015)	TAC 2008 and TAC 2009	ROUGE-2 and ROUGE-SU4	System-NIST, DocHITS, ClusterHITS, System-ceaList1, System-LIPN1 and System-VenessTeam1	On TAC 2008, R-2 Recall=0.06853, Precision=0.07212, F-measure=0.07025, R-SU4 Recall=0.10532, Precision=0.10907, F-measure=0.10714
Learning-based approach for summarizing related sentences (Tzouridis et al. 2014)	RSS feeds of 6 major news sites and news headlines in the field of sports, technology and business	ROUGE-1, ROUGE-2, ROUGE-W, BLEU-1, BLEU-2 and BLEU-3	Random, Shortest, Yen (Yen 1971), Filippova (2010) and Boudin and Morin (2013)	Through a paired t-test at a significance level of 5%, for 100 training instances, R-1=57.66, R-2=43.58, R-W=45.00, B-1=50.39, B-2=47.44 and B-3=44.51
Semantic Role labeling with minimal resources (Kaljahi et al. 2014)	2 datasets described in (Van der Plas et al. 2011) and delivery report of Classic project (van der Plas et al. 2010)	Precision, Recall and F_1-measure	Classic1K, 5K, 1K+5K and SelfT	During Identification, Precision=83.82, Recall=83.66, F_1 = 83.73. During Classification, Precision=67.91, Recall=67.79 and F_1 = 67.85

Table 3 continued

Summarization approach and its author	Dataset used	Evaluation measure	Baseline approaches	Results
Summarizing single documents through nested tree structure (Kikuchi et al. 2014)	RST Discourse Treebank (Carlson et al. 2003)	ROUGE-1	Sentence selection, EDU selection (Hirao et al. 2013), $LEAD_{EDU}$ and $LEAD_{snt}$	For sentence subtree, R-1 = 0.354 and for rooted sentence subtree, R-1 = 0.352
Two-level sparse representation model for summarization of multiple documents (Liu et al. 2015)	DUC 2006 and DUC 2007	ROUGE-1, ROUGE-2 and ROUGE-SU4	Random, Lead (Simon et al. 2007), LSA (Gong and Liu 2001) and DSDR (He et al. 2012)	On DUC 2006, R-1 v0.34439, R-2=0.05122 and R-SU4=0.10717. On DUC 2007, R-1=0.35399, R-2=0.06448 and R-SU4=0.11669
Sparse-coding based reader-aware summarization system for multiple documents (Li et al. 2015a,b)	Own created dataset containing 37 topics and DUC 2006 and DUC 2007	ROUGE-1, ROUGE-2 and ROUGE-SU4	Random, Lead (Wasson 1998), MEAD (Radev et al. 2004a, b), DSDR-non (He et al. 2012), MDS-Sparse+div and MDS-Sparse-div (Liu et al. 2015)	On the own created dataset, R-1=0.438, R-2=0.155 and R-SU4=0.186. On DUC 2006, R-1=0.391, R-2=0.081 and R-SU4=0.136. On DUC 2007, R-1=0.403, R-2=0.092 and R-SU4=0.146
Graph-based Extractive summarization by considering importance, non-redundancy and coherence (Parveen and Strube 2015)	PLOS Medicine dataset and DUC 2002	Human judgements for coherence, ROUGE-SU4, ROUGE-1 and ROUGE-2	Lead (Wasson 1998), Random, MMR (Carbonell and Goldstein 1998), TextRank (Mihalcea and Tarau 2004)	On DUC 2002, R-1 =0.485, R-2=0.230 and R-SU4=0.253. On PLOS medicine dataset with author's abstracts, R-2=0.189 and R-SU4=0.224

Table 3 continued

Summarization approach and its author	Dataset used	Evaluation measure	Baseline approaches	Results
Sparse optimization based compressive document summarization (Yao et al. 2015a, b)	DUC 2006 and DUC 2007	ROUGE-1, ROUGE-2 and ROUGE-SU4	Lead(Wasson 1998), MatrixFacto. (Wang et al. 2008a, b), DsR-Q (Wei et al. 2010), BI-PLSA (Shen et al. 2011), MultiModal. (Wan and Xiao 2009), DSDR (He et al. 2012), Sparsemodel (Liu et al. 2015), PEER 24 and PEER 15 (DUC 2006/2007 participants), CLASSY04 (extractive multi-document summarizer of DUC 2004)	On DUC 2006, R-1 =0.415, R-2=0.094 and R-SU4=0.153. On DUC 2007, R-1 =0.446, R-2=0.124 and R-SU4=0.174
Submodular mixtures based summarization of multi-document hierarchy of topics (Bairi et al. 2015)	About 8000 Wikipedia disambiguation pages	Cluster evaluation metrics such as Jaccard Index, F1-measure and NMI	KM_{docs}, $KMed_{docs}$, $KMed_{topics}$ and LDA_{docs}	In 60% of the disambiguation queries, the proposed approach produces higher JI, F1 and NMI scores than all other baselines
Disaster Summarization through prediction of salient updates (Kedzie et al. 2015)	2014 TREC KBA Stream Corpus (Frank et al. 2012), 2013 and 2014 TREC Temporal Summarization Track data (Aslam et al. 2013)	ROUGE-1, ROUGE-2, Expected Gain and Comprehensiveness	Clustering baselines like Affinity Propagation, Hierarchical Agglomerative Clustering and a Salience baseline such as Rank by Salience	Using R-1, Recall=0.282, Precision=0.344, F-1=0.306 and using R-2, Recall=0.045, Precision=0.056, F-1=0.049. The proposed approach attains the best balance by using Expected Gain and Comprehensiveness

Table 3 continued

Summarization approach and its author	Dataset used	Evaluation measure	Baseline approaches	Results
Summarizing multiple documents through system combination (Hong et al. 2015)	DUC 2001, DUC 2002, DUC 2003, DUC 2004 and TAC 2008, 2009	ROUGE-1 and ROUGE-2	ICSISumm (Gillick et al. 2009), DPP (Kulesza and Taskar 2012), RegSum (Hong and Nenkova 2014), R2N2_ILP (Cao et al. 2015b), PriorSum (Cao et al. 2015c), ClusterCMRW (Wan and Yang 2008; Li et al. 2013; Almeida and Martins 2013; Li et al. 2015a, b)	On DUC 2001, R-1 = 0.3526, R-2 = 0.0788. On DUC 2002, R-1 = 0.3823, R-2 = 0.0946. On DUC 2003, R-1 = 0.3959, R-2 = 0.1018. On DUC 2004, R-1 = 0.3995, R-2 = 0.1048. On TAC 2008, R-1 = 0.3978, R-2 = 0.1208. On TAC 2009, R-1 = 0.4009, R-2 = 0.1200
Phrase-based compressive cross-language summarization (Yao et al. 2015a, b)	DUC 2001 with manual translation of reference summaries into Chinese	ROUGE-1, ROUGE-2, ROUGE-W, ROUGE-L and ROUGE-SU4.	PBES (Phrase-Based Compressive Summarization), Baseline (EN), Baseline (CN), CoRank and Baseline (ENcomp)	With word-based evaluation, R-1 = 0.24917, R-2 = 0.04632, R-W = 0.06252, R-L = 0.13591, R-SU4 = 0.07953

6 Abstractive approaches for text summarization

Abstractive text summarization produces an abstract summary which includes words and phrases different from the ones occurring in the source document. Therefore, abstract is a summary that consists of ideas or concepts taken from the original document but are re-interpreted and shown in a different form. It needs extensive natural language processing. Therefore, it is much more complex than extractive summarization. Table 4 displayed below explains some abstractive text summarization approaches present in the literature.

7 Multilingual approaches for text summarization

When source document is in a number of languages like English, Hindi, Punjabi, Bengali, etc and summary is also generated in these languages, then it is termed as a multi-lingual summarization system. Table 5 below describes a few multilingual approaches present in the literature.

8 Summary evaluation

Evaluation of summary is a very important task in the field of automatic summarization of text. Evaluating the summary besides enhancing development of reusable resources and infrastructure helps in comparing and replicating results and thus, adds competition to improve the results. However, it is practically impossible to manually evaluate multiple documents for obtaining an unbiased view. Therefore, reliable automatic evaluation metrics are required for fast and consistent evaluation. Evaluation of summary is a challenging work too as it is not easy for humans to know what kind of information should be present in the summary. Information changes according to the purpose of the summary and to capture this information automatically, is a difficult task. Figure 6 below describes the taxonomy of summary evaluation measures.

Following are the two ways for determining the performance of text summarization:

- *Extrinsic evaluation*: It determines summary's quality based on how it affects other tasks (Text classification, Information retrieval, Question answering), i.e., a summary is termed as a good summary if it provides help to other tasks. Various methods for extrinsic evaluation are:

  - *Relevance assessment*: Here various methods are used for evaluating a topic's relevance present in the summary or the original document.
  - *Reading comprehension*: It determines whether it is able to answer multiple choice tests after reading the summary.

- *Intrinsic evaluation*: It determines the summary quality on the basis of coverage between the machine-made summary and the human-made summary. Quality or informativeness are the two important aspects on the basis of which a summary is evaluated. Usually the informativeness of a summary is evaluated by comparing it with a human-made summary, i.e., reference summary. There is another paradigm too, i.e. fidelity to the source which checks whether the summary consists of the same or similar content as present in the original document. There is a problem with this method, i.e. how to know which concepts in the document are relevant and which are not.

Table 4 Abstractive text summarization approaches

Technique	Description
Abstractive summarization of more redundant opinions through a graph based approach, Opinosis	Ganesan et al. (2010) presented Opinosis, a new summarization approach that makes use of graphs for generating concise abstractive summaries of highly redundant opinions. Opinosis is highly flexible as it does not require any domain knowledge and it uses shallow NLP. In this approach, firstly a textual graph is made, representing the text to be summarized. Then, for generating candidate abstractive summaries, various sub-paths in the graph are explored and scored by making use of three unique properties of graphs. Evaluation results conclude that summaries generated by Opinosis have reasonable agreement with human summaries. Moreover, readable, concise, well-formed and informative summaries are generated that contain important content. This system is evaluated on reviews of hotels, cars and various products and obtains scores for ROUGE-1 recall as 0.2831, ROUGE-2 recall as 0.0853 and ROUGE-SU4 recall as 0.0851
Abstractive text summarization for Telugu documents	Kallimani et al. (2011) implemented various statistical approaches for abstractive summarization of Telugu documents. The proposed system pre-processes, summarizes and post-processes each document. For summarization, a number of important features are utilized to generate a summary like word clues, keyword extraction, sentence selection, sentence extraction and summary generation. Finally during post-processing, extractive summary is converted to abstractive summary by employing summary refinement and summary rephrasing. The precision obtained for keyword selection over a set of samples is 70 %
Abstractive text summarization through the use of word graphs	Lloret and Palomar (2011a, b) proposed an approach for abstractive text summarization by employing word graphs. This approach compresses and merges information from sentences to form new sentences. Then, an extractive text summarization approach, COMPENDIUM is utilized for determining which of the novel sentences should be selected for forming an abstractive summary. Different approaches are analyzed to discuss issues related to generation of abstracts like how to generate new sentences, order in which relevant content can be selected and length of the sentences. Results show that generation of abstracts is a challenging task. However, experiments prove that by combining extractive and abstractive information, abstracts of better quality can be obtained. ROUGE score of 0.405 is obtained on DUC 2002 by combining both extractive and abstractive approaches

Table 4 continued

Technique	Description
Abstractive text summarization approach through text-to-text generation	Genest and Lapalme (2011) proposed an approach based on the concept of Information Items (INIT) which is the smallest element of coherent information in the text or a sentence. It can be as simple as an entity's characteristic or as complex as the complete description of an event. This approach has four operational steps, i.e. INIT retrieval, sentence generation, sentence selection and summary generation. This approach tries to control the content and structure of the document. Evaluation results on the dataset of TAC 2010 are quite satisfactory. This abstraction system generates summary with pyramid score of 0.315, linguistic quality as 2.174 and overall responsiveness as 2.304
Abstractive text summarization through semantic graph reduction technique	Moawad and Aref (2012) proposed a new method for generating an abstract for single document through a reduction method based on semantic graph. This approach works in three phases: firstly generation of a rich semantic graph from the original documents, secondly reduction of the rich semantic graph thus generated to a highly abstracted graph and finally generation of an abstract. It has been shown by a simulated case study that the given technique minimizes the original text to 50 %
COMPENDIUM, a text summarizer for generation of abstracts of research papers	Lloret and Palomar (2013) proposed a text summarizer, COMPENDIUM, that creates abstracts of biomedical papers. There are two variants of COMPENDIUM, COMPENDIUM$_E$ for generating extracts and COMPENDIUM$_{E-A}$ that contains both abstractive and extractive methods in which after choosing important sentences, information compression and fusion stage is implemented. Then qualitative and quantitative evaluation of this system was done in which it was concluded that COMPENDIUM is suitable for generating summaries as both of its variants are able to select important content from the source document but abstractive-oriented summaries produced by COMPENDIUM$_{E-A}$ are more appropriate from a human perspective. For specialized journal of medicine, ROUGE-1 score for COMPENDIUM$_E$ is 44.02 % whereas for COMPENDIUM$_{E-A}$ is 38.66 %
Semantic Role Labeling (SRL) based abstractive summarization of multiple documents	Khan et al. (2015) proposed an abstractive approach in which summary is not generated by simply selecting sentences from source documents but by semantic representation of the source documents. In this approach, SRL is employed to represent the content of the source document through predicate argument structures. These semantically similar predicate argument structures are clustered by employing semantic similarity measure and then these structures are ranked on the basis of features weighted and optimized by Genetic Algorithm. Experimental results prove that the given approach performs better than the other comparison models and it stood second to the average of human model summaries. On DUC 2002, this abstractive approach has a pyramid score (mean coverage score) of 0.50 and average precision of 0.70

Table 4 continued

Technique	Description
Abstractive summarization of multiple documents through ILP based multi-sentence compression	Banerjee et al. (2015) develops an abstractive summarizer which initially selects the most important document from the multi-document set. Then each sentence from the most important document is used to generate separate clusters. Sentences of other documents that have highest similarity with the cluster sentence are assigned to that cluster. Through a word-graph structure formed from the sentences of each cluster, K-shortest paths are generated. Finally sentences are selected from the set of shortest paths by employing a novel integer linear programming (ILP) problem that maximizes information content and linguistic quality and reducing redundancy in the final summary. Experimental evaluation on DUC 2004 and DUC 2005 datasets show that the ROUGE scores of the proposed system are better than the best extractive summarizer on both the datasets and also this system outperforms an abstractive summarizer based on multi-sentence compression. On DUC 2004, ROUGE-2 score is 0.11992 and ROUGE-SU4 score is 0.14765 and on DUC 2005 ROUGE-L score is 0.35772 and ROUGE-SU4 score is 0.12411
Phrase selection and merging based abstractive summarization of multiple documents	Bing et al. (2015) proposed an abstraction-based summarization system for multiple documents that create new sentences by exploring more fine-grained syntactic units like noun or verb phrases than sentences. Initially a pool of concepts and facts, represented by noun or verb phrases is extracted from the input documents. Then, a salience score is calculated for each phrase by exploiting redundancy of the document content. In order to achieve a global optimum solution, phrases are selected and merged simultaneously leading to the creation of new sentences whose validity is ensured through an integer linear optimization model. Experimental evaluations is carried out on TAC 2011dataset using an automated pyramid evaluation metric. The proposed system scores 0.905 and 0.793 at thresholds 0.6 and 0.65 respectively which is better than the other systems in TAC 2011. Also, this system outperforms the other systems on manual linguistic quality evaluation
Abstractive summarization with a neural attention based model	Rush et al. (2015) proposed a complete data-driven approach for generating abstractive summaries. Based on the recent developments in neural machine translation, a neural attention-model is generated which is combined with a contextual input encoder. This model generates each word of the summary based on the input sentence. Being structurally simple, it can be made to train a large amount of data. It also trains a summarization model for headline generation on Gigaword dataset (Graff et al. 2003) consisting of article pairs. This model is executed on DUC 2004 dataset using ROUGE and it has been shown in the results that this model significantly outperforms several abstractive and extractive baselines. ROUGE scores are: ROUGE-1 = 0.2921, ROUGE-2 = 0.0838 and ROUGE-L = 0.2446

Table 5 Multilingual text summarization approaches

| Technique	Description
Multi-document, multilingual text summarization system, MEAD	Radev et al. (2004a, b) proposed an open source, public domain, extractive, multi-document, multilingual summarization system. Its source and documentation[a] can also be downloaded. This system implements a number of summarization techniques like centroid-based, position-based, query-based, largest common sub-sequence and keywords. Four classifiers are used here: default, lead-based, random and decision tree. MEAD is employed in a number of tasks such as summarization of mobile devices, web pages, novelty detection, etc. It uses two big corpora: SummBank and CSTBank. It supports a number of languages like English and Chinese. It also has an evaluation tool, MEAD Eval in its current version
Multilingual text summarization system using HMSM (Hidden Markov Story Model) based on one story, one flow	Fung and Ngai (2006) proposed a multilingual theme-based summarization approach for multiple documents using stochastic HMSM based on text cohesion. A clustering algorithm, i.e., Modified K-Means (MKM) group several documents into distinct topics (stories). A HMSM is trained for each story from a set of documents in every language through Segmented K-Means (SKM) decoding. SKM helps in classifying sentences into subtopic states through k-means clustering and viterbi decoding. Naïve Bayes Classifier is implemented for the task of summarization that classifies the sentences marked by HMSMs into summary class. Experimental results prove that documents based on one topic (story) have the same flow and also documents in one particular language based on one story have the same flow. Evaluation is done on TDT3 collection dataset using viterbi scoring. This system is implemented on English and Chinese documents. Accuracy of 67.02 % is achieved for Chinese documents and 54.33 % for English documents
Multilingual text summarization through a language independent technique	Patel et al. (2007) proposed a language independent algorithm for generic text summarization for single document. This approach employs structural and statistical features. Being a flexible approach, it requires only a stop words list externally and a stemmer for the respective language in which document needs to be summarized. In this method, a vector is created that represents the theme of the document. The text is partitioned and most important sentences are chosen from each partition. For incomplete sentences, their respective preceding sentences are included to resolve the contextual and semantic gap. Evaluations are performed on English, Hindi, Gujarati and Urdu documents for single document summarization. Results show that for English documents, in 82 % of cases, summaries have a better or equal degree of representativeness as compared to DUC summary. For other language documents, degree of representativeness is more than 80 %

Table 5 continued

Technique	Description
Multilingual news summarization system, NewsGist based on statistical technique	Kabadjov et al. (2010) presented a multilingual news summarization system, NewsGist for multiple documents based on SVD (Singular Value Decomposition). It is developed for EMM (Europe Media Monitor) that collects a large number of news articles in various languages from many on-line news sources and groups them into important news clusters. This summarizer then generates summaries for each distinct news cluster. The task of summarization has three phases: interpretation, transformation and generation. In interpretation phase, term-by-sentence matrix is developed for a collection of documents. Then SVD is applied to term-by-sentence matrix in the transformation phase. Finally the summary is generated by selecting only relevant sentences. This summarizer is used by EMM for some languages like English, German, French, etc
Arabic text summarizer using RST and scoring of sentences	Azmi and Al-Thanyyan (2012) presented a text summarization system based on extractive technique for Arabic language in which no machine learning is used and the user can finally restrict the length of the summary. This algorithm has two phases: (a) Phase 1: using RST, a primary summary is created and (b) Phase 2: in the primary summary, score of each sentence is computed. Then, sentences are chosen for the final summary while taking into consideration that summary's total score is maximum while summary is within the size limit fixed by the user. RST describes the text and their coherence. All the text units are connected together to create a rhetorical structure that is usually depicted by trees. The evaluation is done on two Saudi newspapers: Ar-Riyadh and Al-Jazirah. The summary of size 31 % generated by this system has precision as 0.66, recall as 0.70 and F-measure as 0.67. Implementation of this system can also be done for languages like Farsi and Urdu
Multilingual summarizer for Hindi and Punjabi documents using a hybrid algorithm	Gupta (2013) proposed a hybrid algorithm for summarizing multilingual text documents belonging to Hindi and Punjabi. This method employs features of Hindi text summarization system suggested by CDAC Noida as well as Punjabi text summarization system (Gupta and Lehal 2010). This technique implements these nine features: key phrase, font, nouns and verbs, position, cue-phrase, negative keywords, named entity, relative length and numerical data. Machine learning based mathematical regression is employed to compute feature weights from training set of documents. Then score of each sentence in the test data is computed for each of the nine features. At 30 % compression rate, highly scored sentences are chosen to form the summary. This method performs better at a compression rate of 30 % for both extrinsic and intrinsic summary evaluation measures and obtains a good F-score of 92.56 %

Table 5 continued

Technique	Description
An approach for multilingual text summarization using distributed representations of word and mRMR discriminant analysis	Oufaida et al. (2015) proposed a multilingual text summarization system that selects important sentences from single as well as multiple documents with mRMR approach (minimum redundancy and maximum relevance). This approach follows a two-step summarization process: (a) Sentences are initially clustered by k-Medoids algorithm with the help of semantic content present in word representations and (b) Initially terms and then sentences are scored through informativeness of words using discriminant analysis approach. A new metric for sentence similarity is proposed to find best similarity between words in two sentences. Based on the required size of the summary, a new two speed sentence extraction algorithm is proposed. This system is implemented on three languages: English, French and Arabic. Evaluation is done on TAC Multiling 2011 dataset by using two evaluation metrics: ROUGE and MeMoG (Giannakopoulos et al. 2008). This system produces comparable results for English (MeMoG score = 0.155) and French (MeMoG score = 0.164) but Arabic results need to be improved further (MeMoG score = 0.117)

^a http://www.summarization.com/mead
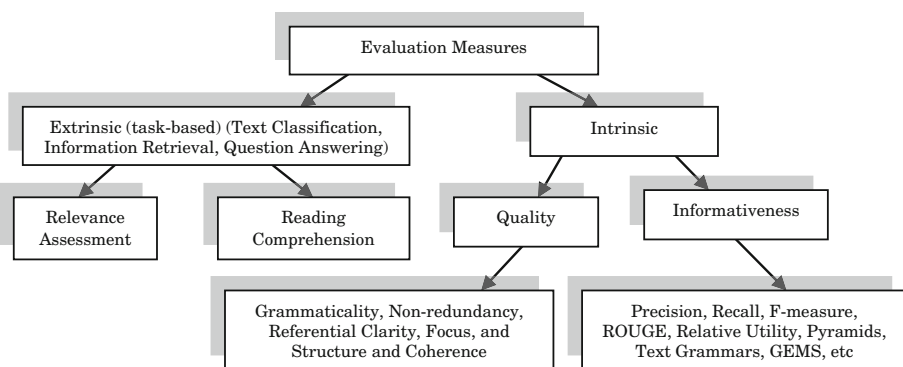
Fig. 6 Taxonomy of summary evaluation measures

8.1 Informativeness evaluation

Some of the methods for informativeness evaluation are Relative utility, Factoid score, Pyramid method, ROUGE (Lin 2004), etc. ROUGE counts the number of units common between a particular summary and a collection of reference summaries. Thus, it helps to automatically evaluate the summary. ROUGE includes five measures like ROUGE-N, ROUGE-L, ROUGE-W, ROUGE-S and ROUGE-SU. Their explanation is as follows:

- ROUGE-N measures the N-gram units common between a particular summary and a collection of reference summaries where N determines the N-gram's length. E.g., ROUGE-1 for unigrams and ROUGE-2 for bi-grams.
- ROUGE-L computes Longest Common Subsequence (LCS) metric. LCS is the maximum size of common subsequence for two given sequences X and Y. ROUGE-L calculates ratio between size of two summaries' LCS and size of reference summary.
- ROUGE-W is the weighted longest common subsequence metric. It's the improvement over the simple LCS approach. ROUGE-W prefers LCS with successive common units. It can be computed efficiently using dynamic programming.
- ROUGE-S (Skip-Bigram co-occurrence statistics) evaluates the proportion of skip bigrams common between a particular summary and a collection of reference summaries. Any word pair in the sentence order with random gaps is the skip bigrams.
- ROUGE-SU is the weighted average between ROUGE-S and ROUGE-1 and it extends ROUGE-S with counting unit as unigram. Actually this is an improvement over ROUGE-S.

For intrinsically evaluating the summary, other popular metrics are precision, recall and f-measure. They are required to predict coverage between human-made summary and automatically generated machine-made summaries. With the help of above metrics, it is also feasible that two summaries generate different evaluation results even being identically good. These metrics are explained below:

- *Precision*: It determines what fraction of the sentences chosen by the humans and selected by the system are correct.
- *Recall*: It determines what proportion of the sentences chosen by humans are even recognized by the machine.
- *F-measure*: It is computed by combining recall and precision.

Explanation of some other methods for informativeness evaluation is given below:

- *Relative utility* (Radev and Tam 2003): In this metric, judges assign a score between 0 and 10 to each sentence in the input document according to its relevance. The highly scored sentences are considered more appropriate for the summary.
- *Text grammars*: This method helps to evaluate text summaries. Structure of valid text is expressed in a formal manner through this method.
- *Factoid Score* (Teufel and Halteren 2004): Evaluation of automatic summaries is done with respect to factoids (these are atomic units of information that are used to express the meaning of a sentence). Different reference summaries are used as gold standards and common information is measured among them.
- *BE (Basic Elements)*: Here a sentence is segmented into very tiny units of content, known as BE, that are expressed as words' triplets (head |modifier| relation), containing a head, modifier or argument along with the relationship of modifier to the head. Aim of this method is to match different equivalent expressions with more flexibility.
- *Pyramid Method*: It searches for information with same meaning known as Summary Content Units (SCU) in various human-made summaries. A weight is assigned to each SCU corresponding to the number of human assessments which identify the same content. These weights have a particular distribution that distinguishes relevant information from less relevant one.
- *AutoSummENG (Automatic Summary Evaluation based on N-gram Graphs)* (Giannakoloulos et al. 2008b): It is having high correlation with human judgement and it is an automatic method. For comparing the summaries, n-gram character graphs are ini-
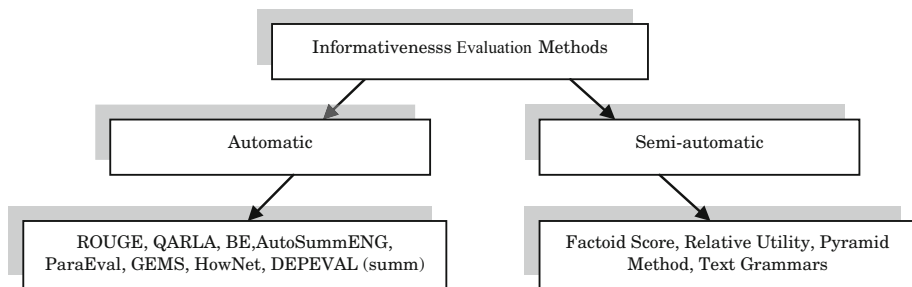
Fig. 7 Automatic and semi-automatic methods for informativeness evaluation

tially built and their representations are compared to obtain some sort of similarity among the graphs. This approach is language-independent.

- *QARLA*: Amigó et al. (2005) suggested this evaluation framework. Given some automatic and reference summaries and some similarity metrics, this approach provides some measures like QUEEN, (which evaluates the quality of a machine-generated summary) KING (which evaluates a similarity metric's quality) and JACK (that is used for estimating the reliability of machine-generated summaries). This framework uses a total of 59 distinct similarity metrics like precision, recall, frequency and sentence length and metrics for grammatical distribution.
- *ParaEval*: It is proposed in Zhou et al. (2006). It is used for detecting paraphrase matching. Process of paraphrase detection occurs in three steps. Initially paraphrases composed of multiple words are searched between phrases in the reference and automatic summaries. In second step, this method tries to look for synonyms between single words for those unmatched fragments. Finally, if no synonym is found between single words, then simple lexical matching is done.
- *DEPEVAL (summ)*: It is a dependency-based metric, suggested by Owczarzak (2009). It has a concept similar to Basic Elements (BE) except that parsers are used here and Minipar is used in BE. Here dependency triples are selected from automatic and reference summaries and are then compared with one another.

Some of the above informativeness evaluation methods are automatic, which do not require human annotations while others are semi-automatic, which require some sort of human annotations. Factoid score, Relative utility, Pyramid method and Text grammars are semi-automatic while rest of the others are automatic. List of automatic and semi-automatic methods is displayed above in Fig. 7.

8.2 Quality evaluation

Here linguistic aspects of the summary are considered. In the conferences of DUC and TAC, five questions based on linguistic quality are employed for evaluating summaries like non-redundancy, focus, grammaticality, referential clarity, and structure and Coherence that do not need to be compared against the reference summary. Expert human assessors evaluate the summary manually by assigning a score to the summary corresponding to five-point scale on the basis of its quality.

Text quality of summary can also be assessed by analyzing different factors for readability (Pitler and Nenkova 2008). Text quality is analyzed through different criterions like vocabulary, syntax or discourse so that correlation can be estimated between these factors

and already obtained human readability ratings. Vocabulary is expressed by unigrams and syntax by features like average number of verb-phrases or noun-phrases. Other quality evaluation paradigms are local coherence (Barzilay and Lapata 2005), centering theory (Grosz et al. 1995), syntactic and semantic models and grammaticality of a grammar (Vadlapudi and Katragadda 2010).

8.3 Asiya, an evaluation toolkit

Asiya is an automatic machine translation evaluation and meta-evaluation toolkit. It has a collection of metrics and meta-metrics. It serves as an interface to a compiled set of evaluation and meta-evaluation methods. In the metric repository, there are current versions of most popular metrics grouped in three distinct linguistic levels, i.e., syntactic, lexical and semantic and on the basis of various similarity metrics like recall, precision, overlap, etc. Under lexical similarity, metrics are BLEU, NIST, GTM, METEOR, ROUGE, TERp and O_l. Syntactic similarity consists of metrics like shallow parsing, dependency parsing and constituency parsing. Semantic similarity includes metrics like named entities, semantic roles and discourse representations. Asiya works over fixed sets of translation test cases that are pre-defined test suites. Meta-metric repository contains KING (Amigó et al. 2005) and ORANGE (Lin and Och 2004a) which are measures depending on human acceptability, i.e., correlation with human likeness and human assessments.

8.4 Text summarization evaluation programs

The first conference where automatic summarization systems were evaluated was held at the end of 90's and it was named as SUMMAC (TIPSTER Text Summarization Evaluation) (Mani and Maybury 1999)[2] and here text summaries were evaluated using two extrinsic and one intrinsic method. Single-document query-based summaries of newswire documents were evaluated in this evaluation program.

Another evaluation program, NTCIR (National Institute for Informatics Test Collection for IR) [3] formed a series of three Text Summarization Challenges (TSC) workshops- TSC in 2001, TSC2 in 2002 and TSC3 in 2003 which incorporated Japanese summarization tasks and the evaluation was done using both extrinsic and intrinsic evaluation methods.

The other important conference for text summarization is DUC (Document Understanding Conferences)[4] that was held every year from 2001 to 2007 in two phases, such that first phase consisted of DUC 2001–DUC 2004 and second phase consisted of editions from DUC 2005 onwards after incorporating a revision. All the editions of this conference contained documents of newswire domain. During these DUC conferences, the summarization systems improved and various summary evaluation methods were proposed to meet new challenges and needs of text summarization systems. Change occurred from an entire manual evaluation in which SEE[5] evaluation environment was used to a complete automatic evaluation package ROUGE (Lin 2004) and Basic Elements (Hovy et al. (2006)). Initially in DUC conferences like DUC 2001 and DUC 2002, tasks involved generic summarization of single and multiple documents and later on extended to query-based summarization of multiple documents in DUC 2003. In DUC 2004, topic based single and multi-document cross-lingual summaries

---

were evaluated. In DUC 2005 and DUC 2006, multi-document, query-based summaries were evaluated whereas in DUC 2007, multi-document, update, query-based summaries were evaluated. These conferences besides evaluating and comparing automatic text summarization systems, also provides standard corpora of documents and gold summaries and can be availed on demand.

But after 2007, DUC conferences were not organized as they got included in Text Analysis Conference (TAC) [6] in which summarization tracks are present. TAC is a series of evaluation workshops which are held for promoting research in the area of Natural Language Processing and other similar fields. It acts as a forum for organizations for sharing their results as it provides large test collections and common evaluation techniques. TAC consists of tracks which are sets of tasks, each focused on a specific sub-problem of NLP. TAC tracks consist of end-user tasks as well as component evaluations within the context of end-user tasks. A mailing list is contained in each track for discussing the task details present in the track in the latest TAC cycle.

The TAC QA track evolved from the TREC QA track. Summarization track helps in developing systems for generating short, coherent text summaries. TAC 2008 QA track helps to answer short series of opinion questions in which each series of 2–4 questions is about a particular target. 2008 summarization track consists of two tasks: Update task and Opinion pilot. Update summarization's task is to write a short summary (around 100 words) from a collection of news articles, assuming that the user has already gone through a collection of previous articles. The opinion pilot's task is to write summaries of opinions from blogs. 2009 Summarization track has two tasks: Update summarization which is the same as in 2008 summarization track and Automatically Evaluating Summaries of Peers (AESOP). AESOP computes summary's score with respect to a particular metric that is related to the summary's content like overall responsiveness and pyramid scores. AESOP is a new task that was introduced in 2009 which enhanced the basic summarization task by constructing a collection of automatic summarization tools that help in developing summarization systems.

2010 summarization track has two tasks: guided summarization and AESOP. Guided summarization's task is the generation of a 100 word summary from a collection of 10 news articles belonging to a specific topic such that each topic belongs to a category previously defined. This task helps to promote a greater linguistic (semantic) analysis of the original documents despite of depending solely on the document word frequencies for selecting relevant concepts. AESOP task is the same as in 2009 summarization track. 2011 summarization track consists of three tasks: guided summarization, AESOP and Multiling (i.e. Multilingual) pilot. Guided summarization and AESOP tasks are the same as in 2010 summarization track. Multiling pilot task performs summarization using multi-lingual algorithms. TAC 2012 emphasizes on Knowledge Based Population (KBP). KBP encourages research in automated systems which search for information related to named entities as obtained from a big corpus and add this information into a Knowledge Base (KB). TAC KBP 2012 track consists of tasks in three areas: entity linking, slot filling and cold start KBP. TAC KBP 2013 track consists of the following tasks: entity linking, English slot filling, temporal slot filling, cross-lingual Spanish slot filling, sentiment slot filling, slot filler validation and cold start KBP. TAC summarization track 2014 deals with biomedical summarization. KBP 2014 track consists of the followings tasks: cold start KBP, entity linking, slot filling, slot filler validation, sentiment and event. TAC 2015 summarization track consists of cold start KBP track, tri-lingual Entity Discovery and Linking track (EDL), event track and validation/ensembling track. Table 6 illustrates automatic text summarization evaluation conferences along with their respective

---

Table 6 Text summarization evaluation conferences along with their respective summarization task features

Conference	Summarization task features
SUMMAC	Single-document, query-based summarization of newswire documents
TSC (NTCIR)	Query-based, generic summarization of newswire documents
TSC2 (NTCIR)	Single and multi-document, generic summarization of newswire documents
TSC3 (NTCIR)	Multi-document, generic summarization of newswire documents
DUC-01	Single and multi-document, generic summarization of newswire documents
DUC-02	Single and multi-document, generic summarization of newswire documents
DUC-03	Multi-document, query-based summarization of newswire documents
DUC-04	Single and multi-document, cross-lingual, topic-oriented summarization of newswire documents
DUC-05	Multi-document, query-based summarization of newswire documents
DUC-06	Multi-document, query-based summarization of newswire documents
DUC-07	Multi-document, query-based, update summarization of newswire documents
TAC-08	Multi-document, update, query-based, sentiment-based summarization of newswire documents and blogs
TAC-09	Multi-document, update, query-based summarization of newswire documents, evaluation
TAC-10	Multi-document, guided, query-based summarization of newswire documents, evaluation
TAC-11	Multi-document, guided, query-based, multi-lingual summarization of newswire documents, evaluation
TAC-12	Multi-document, guided, query-based, entity-linking, slot filling, cold start KBP, multi-lingual, cross-lingual summarization of newswire documents, evaluation
TAC-13	Multi-document, guided, query-based, temporal-based, sentiment-based, entity-linking, slot filling, cold start KBP, multi-lingual, cross-lingual summarization of newswire documents, evaluation
TAC-14	Multi-document, guided, query-based, event-based, temporal, sentiment-based, entity-linking, slot filling, cold start KBP, multi-lingual, cross-lingual summarization of bio-medical documents, evaluation
TAC-15	Multi-document, guided, query-based, event-based, validation-based, entity-linking, slot filling, cold start KBP, multi-lingual, cross-lingual summarization of newswire documents, evaluation

summarization task features. The tasks that are underlined are new for that conference with respect to the previous conferences.

9 Evaluation results

In this section, evaluation results of the extractive summarization approaches surveyed in this paper are discussed. By focusing on experimental works, performance of various text summarization methods is reported on shared DUC datasets using an automatic evaluation framework, ROUGE which is DUC's official evaluation metric for summarization of text. Shared datasets like DUC 2002, DUC 2004 and DUC 2007 are chosen for discussing the evaluation results. For comparing the performance of different text summarization methods, various variants of ROUGE are used like ROUGE-1 (unigram overlap), ROUGE-2 (bigram

overlap) and ROUGE-SU4 (skip bigram with unigram as counting unit). The summaries generated by these summarization approaches have comparable length (200–250 words) to ensure a fair evaluation. Table 7 below gives a brief description of text summarization approaches surveyed in this paper as well as other techniques used for comparison in the evaluation process.

Table 7 Text summarization approaches used in the evaluation process

Technique	Description
OCDsum-SaDE (Alguliev et al. 2013)	This is an optimization approach for generic summarization of documents. This approach deals with content coverage and redundancy at the same time. This approach uses an algorithm named as self-adaptive DE (Differential Evolution) for solving the problem of optimization and it uses crossover, mutation and selection operators
UnifiedRank (Wan 2010)	This is a graph based approach in which summarization of both single and multiple documents is done simultaneously. This method studies the mutual influences between the two tasks
BSTM (Wang et al. 2009)	Bayesian Sentence based Topic Model (BSTM) employs both term-sentence and term document associations for summarizing multiple documents
FGB (Wang et al. 2011)	Factorization with Given Bases (FGB) is a language model where sentence bases are the given bases and it utilizes document-term and sentence term matrices. This approach clusters and summarizes the documents simultaneously
MA-SingleDocSum (Mendoza et al. 2014)	This is an extractive, generic summarization method for single-document that uses generic operators and guided local search. This method uses a memetic algorithm which has combined the population based search of evolutionary algorithm with a guided local search strategy
DE (Aliguliyev 2009)	A method named as differential evolution is implemented in this approach which optimizes the allocation of each sentence to a group. Sentence selection for the summary depends on the measure of centrality of each sentence with respect to its corresponding group that is measured through Normalized Google Distance
NetSum (Svore et al. 2007)	This is a neural network based summarization approach. In this approach, RankNet learning algorithm is implemented to train a pair-based sentence ranker with the help of which a score is assigned to each sentence in the document and then highly scored sentences are selected
NMF (Lee et al. 2009)	This is an unsupervised summarization approach for generic documents by using Non-negative Matrix Factorization (NMF). In this method, components of semantic feature vectors entirely contain non-negative values and they are also so sparse that semantic features can be interpreted very well
EventGraph-based approach (Glavaš and Šnajder 2014)	This is an event-based summarization and information retrieval model that depends on event extraction at sentence-level. Circumstances of an event are narrated in text by event-mentions which are used to represent real-world events
Progressive approach (Ouyang et al. 2013)	This is a new progressive method for generating summary by selecting "novel and salient" sentences. Only uncovered concepts are examined here for saliency estimation

Table 7 continued

Technique	Description
TAOS (Fang et al. 2015)	Topic Aspect-Oriented Summarization (TAOS) is based on topic factors. These topic factors are various features that describe topics such as capital words are used to represent entity. Various topics can have various aspects and various preferences of features are used to represent various aspects
Sub-SVM (Sipos et al. 2012)	This is a supervised learning approach that is applicable to all sub-modular scoring functions ranging from pair-wise similarity models to coverage based approaches
BudSub (Lin and Bilmes 2010)	This is an unsupervised summarization approach that maximizes sub-modular functions with a constraint on budget through a greedy algorithm. Budget here refers to the length of the summary
ItemSum (Baralis et al. 2012)	This is a multi-document summarization system that depends on an itemset model consisting of frequent itemsets. It selects most relevant and non-redundant sentences for the summary that covers itemset based model in the best way using a sentence relevance score based on TF–IDF statistics
MCKP (Takamura and Okumura 2009)	This approach considers summarization of text as a maximum coverage problem. Some decoding algorithms have been used to summarize text for solving MCKP like stack decoding, greedy algorithm with performance guarantee, branch and bound method and linear relaxation problem with randomized decoding
Ranking-based MMR (Yang et al. 2014)	This is a ranking-based sentence clustering framework in which a term is treated like a text object which is independent rather than the feature of a sentence. Clusters contain highly related sentences. Various topic themes are discovered and clusters are based on these themes
OTS[a] (2011)	OTS is an open source tool for summarizing texts. This is a library as well as a command line tool. AbiWord and KWord are the word processors that can link to the library and summarize documents while the command line tool helps to summarize text on the console. This system is multilingual and supports more than 25 languages which are configured in XML files
MCMR (B&B) (Alguliev et al. 2011)	Maximum Coverage and Minimum Redundancy (MCMR) approach is an unsupervised generic text summarization model that considers summarization of text as an Integer Linear Programming problem (ILP). Branch & Bound Algorithm (B&B) is an optimization algorithm that is used to solve ILP problem which is an NP-hard problem
MCMR (PSO) (Alguliev et al. 2011)	Maximum Coverage and Minimum Redundancy (MCMR) approach is an unsupervised generic text summarization model that considers summarization of text as an Integer Linear Programming problem (ILP). Particle Swarm Optimization (PSO) is an optimization technique used for solving ILP problem

Table 7 continued

Technique	Description
AdaSum (Zhang et al. 2008b)	AdaSum is an adaptive model for topic-based multi-document summarization that can optimize topic representations as well as generate effective summaries
Uni + Max (Ouyang et al. 2011)	This is a unigram based approach with maximum scoring function. It is an approach based on Support Vector Regression (SVR) that ranks and selects important sentences by employing a set of predefined features
PNR^2 (Wenjie et al. 2008)	PNR^2 (Ranking with Positive and Negative Reinforcement) is a graph based sentence ranking approach for update summarization. During the process of ranking, it considers both positive and negative mutual reinforcement
MDS-Sparse-div (Liu et al. 2015)	This is a two-level sparse representation model for multi document summarization that employs document reconstruction and is based on important properties of an ideal reconstructable summary: coverage and sparsity and it doesn't consider diversity
Sum_Sparse (Li et al. 2015a, b)	This is a reader-aware summarization system for multiple documents (RA-MDS) based on sparse-coding technique that generates summaries not only from the reports of the events but also considers the reader comments at the same time
Sum_Coh (Parveen and Strube 2015)	This is graph-based unsupervised technique for extractive summarization of single documents which considers three important properties of summarization, i.e. importance, non-redundancy and local coherence
SpOpt-comp (Yao et al. 2015a, b)	This is sparse optimization based extractive document summarization which has a decomposable convex objective function that is solved by an efficient ADMM (alternating direction method of multipliers) algorithm
SumCombine (Hong et al. 2015)	This is a multi-document summarization approach in which summaries generated from different systems are combined

[a] http://libots.sourceforge.net/

Figure 8 below shows the comparison of various text summarization methods on DUC 2002 using ROUGE-1 and ROUGE-2. As can be seen below in Table 8, ROUGE-1 score varies from 0.3823 for SumCombine (Hong et al. 2015) to 0.4990 for OCDsum-SaDE (Alguliev et al. 2013). ROUGE-2 score varies from 0.0946 for SumCombine (Hong et al. 2015) to 0.2548 for OCDsum-SaDE (Alguliev et al. 2013). Best scores of ROUGE-1 and ROUGE-2 are obtained by OCDsum-SaDE because this approach employs an optimization algorithm named as self-adaptive DE (Differential Evolution) and it reduces redundancy in the summaries as well as selects important sentences from the documents, covering the relevant content of the original document. Lowest score of ROUGE-1 and ROUGE-2 are obtained by SumCombine.

Figure 9 below shows the comparison of different text summarization methods on DUC 2004 using ROUGE-1. Table 9 displayed below shows that ROUGE-1 score ranges from
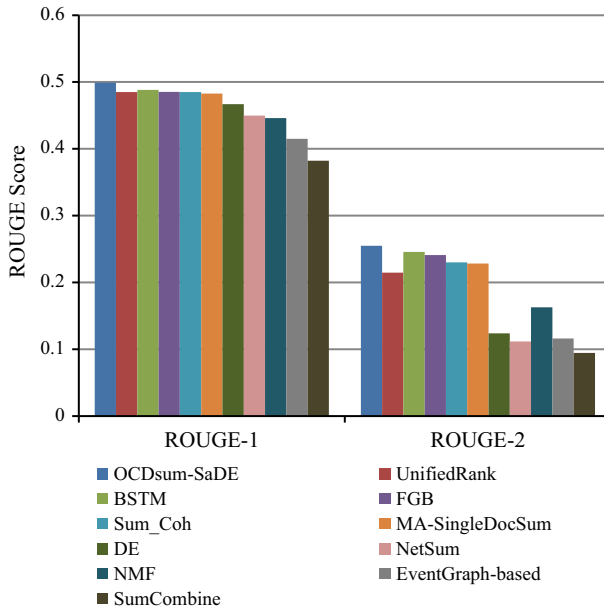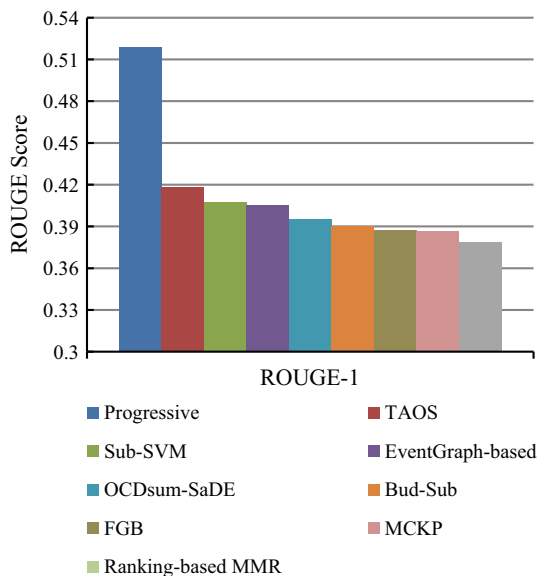
Fig. 8 Comparison of text summarization methods on DUC 2002

Table 8 ROUGE score of the text summarization methods on DUC 2002 dataset

Methods	ROUGE-1 (with rank)	ROUGE-2 (with rank)
OCDsum-SaDE	0.4990 (1)	0.2548 (1)
UnifiedRank	0.4849 (2)	0.2146 (6)
BSTM	0.4881 (3)	0.2457 (2)
FGB	0.4851 (4)	0.2410 (3)
Sum_Coh	0.4850 (5)	0.2300 (4)
MA-SingleDocSum	0.4828 (6)	0.2284 (5)
DE	0.4669 (7)	0.1237 (8)
NetSum	0.4496 (8)	0.1117 (10)
NMF	0.4459 (9)	0.1628 (7)
EventGraph-based	0.4150 (10)	0.1160 (9)
SumCombine	0.3823 (11)	0.0946 (11)

0.3788 for Ranking-based MMR approach (Yang et al. 2014) to 0.5190 for Progressive approach (Ouyang et al. 2013). Highest ROUGE-1 score is obtained by Progressive approach because a conditional saliency measure of sentences is used here which discovers subsuming relationship among sentences despite of general saliency measures employed in maximum of the prevailing approaches. Therefore, relevant general concepts help to explore relevant supporting concepts. Lowest score of ROUGE-1 is obtained by Ranking-based MMR approach.
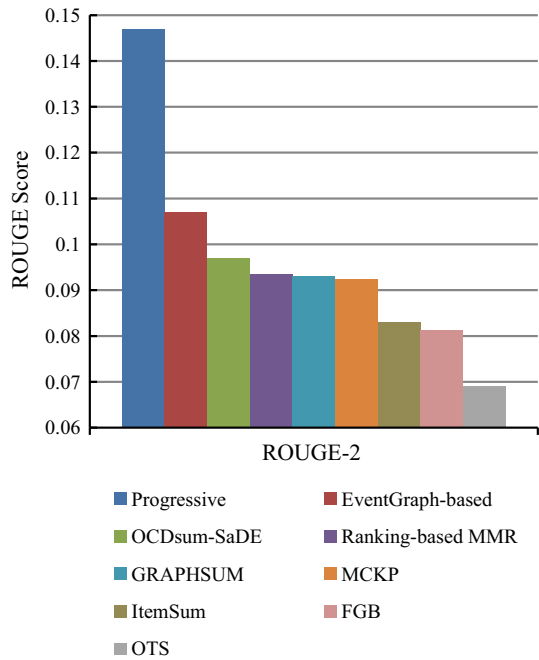
Figure 10 below shows the comparison of different text summarization methods on DUC 2004 using ROUGE-2. As can be seen below in Table 10, ROUGE-2 score varies from 0.0690 for OTS to 0.1470 for Progressive approach (Ouyang et al. 2013). Highest ROUGE-2 score is obtained by Progressive approach because a conditional saliency measure of sentences

Fig. 9 Comparison of the text summarization methods using ROUGE-1 on DUC 2004

Table 9 ROUGE-1 score of the text summarization methods on DUC 2004 dataset

Methods	ROUGE-1 (with rank)
Progressive	0.5190 (1)
TAOS	0.4185 (2)
Sub-SVM	0.4074 (3)
EventGraph-based	0.4050 (4)
OCDsum-SaDE	0.3954 (5)
Bud-Sub	0.3901 (6)
FGB	0.3872 (7)
MCKP	0.3864 (8)
Ranking-based MMR	0.3788 (9)

is used here which discovers subsuming relationship among sentences despite of general saliency measures employed in maximum of the prevailing approaches. Therefore, relevant general concepts help to explore relevant supporting concepts. Lowest score of ROUGE-2 is obtained by OTS.

Figure 11 below shows the comparison of different text summarization methods on DUC 2007 using ROUGE-2 and ROUGE-SU4. Table 11 displayed below shows that ROUGE-2 score ranges from 0.0645 for MDS-Sparse-div (Liu et al. 2015) to 0.1262 for Ranking-based MMR (Yang et al. 2014). ROUGE-SU4 score varies from 0.1167 for MDS-Sparse-div (Liu et al. 2015) to 0.1780 for Ranking-based MMR (Yang et al. 2014). Best score of ROUGE-2 and ROUGE-SU4 are obtained by Ranking-based MMR approach because this approach generates high quality sentence clusters based on theme and it uses a modified MMR-like approach to control redundancy in summarization of multiple documents. Lowest score of ROUGE-2 and ROUGE-SU4 are obtained by MDS-Sparse-div.
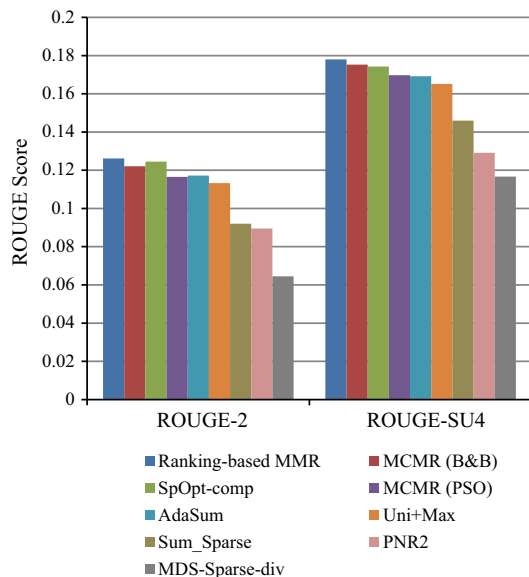
Fig. 10 Comparison of the text summarization methods using ROUGE-2 on DUC 2004

Legend:
- Progressive
- EventGraph-based
- OCDsum-SaDE
- Ranking-based MMR
- GRAPHSUM
- MCKP
- ItemSum
- FGB
- OTS

Table 10 ROUGE-2 score of the text summarization methods on DUC 2004 dataset

Methods	ROUGE-2 (with rank)
Progressive	0.1470 (1)
EventGraph-based	0.1070 (2)
OCDsum-SaDE	0.0969 (3)
Ranking-based MMR	0.0936 (4)
GRAPHSUM	0.0930 (5)
MCKP	0.0924 (6)
ItemSum	0.0830 (7)
FGB	0.0812 (8)
OTS	0.0690 (9)

10 Future directions in text summarization

There has been tremendous research in the field of text summarization over the past fifty years. Novel approaches have been developed that incorporate linguistic aspects into the summary so now the summary is not just the simple concatenation of sentences. This research field is improving continuously, meeting new needs of users and facing a number of challenges. Therefore, in this section, focus is emphasized on the important issues arising in this field of research that needs to be addressed by the research community.

Existing text summarization methods are updating with time like new machine learning algorithms are employed to build text summarization systems. But there is not much change in the features (term frequency, position, etc) required to extract important sentences. There-

Fig. 11 Comparison of the text summarization methods on DUC 2007

Table 11 ROUGE score of the text summarization methods on DUC 2007 dataset

Technique	ROUGE-2	ROUGE-SU4
Ranking-based MMR	0.1262 (1)	0.1780 (1)
MCMR (B&B)	0.1221 (3)	0.1753 (2)
SpOpt-comp	0.1245 (2)	0.1743 (3)
MCMR (PSO)	0.1165 (5)	0.1697 (4)
AdaSum	0.1172 (4)	0.1692 (5)
Uni+Max	0.1133 (6)	0.1652 (6)
Sum_Sparse	0.0920 (7)	0.1460 (7)
PNR^2	0.0895 (8)	0.1291 (8)
MDS-Sparse-div	0.0645 (9)	0.1167 (9)

fore, some new features for words and sentences need to be discovered which can extract semantically important sentences from the document.

There is change in the type of summaries to adapt to changing user requirements. Initially generic single document summaries were generated but now because of availability of large amount of data in different formats and different languages and due to fast development of technology, multi-document, multi-lingual, multimedia summaries have gained popularity. This is also evident from the evaluation programs which are now working on new types of summarization tracks. Summaries with specified focus like sentiment-based, personalized summaries, etc are also being generated. But how such information can be presented is another important issue. At present most of the systems deal with textual input and output. New approaches can be proposed in which input can be in the form of meetings, videos, etc and output in a format, other than text. Some other systems can be developed in which input is in the form of text and output can be represented through statistics, tables, graphics, visual-rating scales, etc that allows visualization of the results and users can access the required content in less time.

Many new approaches have been proposed that deal with linguistic features and have improved the quality of summaries. But summarization systems based on linguistic approaches require more processor and memory space as they need more linguistic knowledge and difficult linguistic techniques. Moreover, there is an additional complexity in employing linguistic resources (Context Vector Space, Lexical Chain, WordNet, etc) and linguistic analysis tools (discourse parser) of good quality as there is a scarcity of different language resources. Therefore, there is a need to develop statistical based efficient summarization systems that can summarize texts of all languages and generate a summary whose quality matches that of a human summary.

Apart from concatenating the sentences, content in the summary needs to be coherent. Therefore, abstractive or hybrid approach needs to be improved more. With hybrid techniques, important information can be selected, merged, compressed or some information can be deleted to obtain new summary information. Hybrid approach can be developed to produce a good quality summary by combining extractive and abstractive techniques together. Research is also going on to generate abstracts so that the machine generated summaries match closely to the human-written ones.

Another big challenge is the evaluation process. This paper discussed both the types of evaluation methods, intrinsic as well as extrinsic. Most of the evaluation is intrinsic in nature which is further categorized into informativeness and quality evaluation and it is carried out through recent methods and tools. Majority of the recent tools assess the information present in the summary and very few methods try assessing the summary quality. New approaches are being developed to automate the quality evaluation process which is an entire manual process performed by expert judges. Generally, available intrinsic evaluation methods focus on the vocabulary common between a machine-generated and reference summary. Research can be carried out in intrinsic evaluation, thus devising new ways to evaluate the summary on the basis of information it contains and its presentation. Evaluation process is highly subjective. Firstly, a good criterion needs to be defined so that it is clear to the system that what is important and what is not. It is also not known whether this process can be sufficiently automated. Similarly, quality evaluation of summary is also highly subjective, since it is performed manually by expert judges. There are some metrics for quality assessment also like grammaticality, coherence, etc but different results are obtained when same summary is evaluated by two experts.

Text summarization is more than fifty years old and research community is greatly interested in this field so they keep on improving existing text summarization approaches or develop novel summarization approaches to generate summaries of higher quality. But still performance of text summarization is moderate and summaries generated are not so perfect. Therefore, this system can be made more intelligent by combining with other systems so that the combined system can perform better.

11 Conclusion

Text summarization is an interesting research field and it has a wide range of applications. The objective of this paper is to make researchers familiar with some important information related to the past of text summarization, current state-of-the-art and possibilities for the future. The survey carried out in this paper would serve as a good starting point for the novice researchers to gain insight into the main issues related to text summarization. In this paper, classification of well known extractive approaches of text summarization is done into

different categories. Novel type of summaries that have emerged recently is discussed. Summary evaluation is another challenging issue in this research field. Therefore, both methods of summary evaluation are discussed in detail, i.e., intrinsic as well as extrinsic along with text summarization evaluation programs that have occurred till date. Especially, more focus is emphasized on recent extractive approaches of text summarization developed in the last decade. A list of pros and cons of these approaches along with the need of each technique will definitely help the readers know the usefulness of each technique. A brief description of a few abstractive and multilingual techniques is also provided. In addition, all these techniques have been compared in a tabular form, providing some more useful information about these approaches. Further, evaluation results are presented on some shared DUC datasets. Evaluation results show that among the recent text summarization approaches surveyed in this paper, best scores of ROUGE-1 and ROUGE-2 are procured on DUC 2002 dataset through an optimization based approach, OCDsum-SaDE (Alguliev et al. 2013). Whereas on DUC 2004, Progressive approach (Ouyang et al. 2013) has produced highest ROUGE-1 and ROUGE-2 scores. Moreover, a clustering-based approach, Ranking-based MMR (Yang et al. 2014) has shown best ROUGE-2 and ROUGE-SU4 scores on DUC 2007. Finally, some good future directions are provided to the researchers that will help them in improving summary generation techniques so that this research field progresses continuously.

References

Abuobieda A, Salim N, Albaham AT, Osman AH, Kumar YJ (2012) Text summarization features selection method using pseudo genetic-based model. In: International conference on information retrieval knowledge management, pp 193–197

Aliguliyev RM (2009) A new sentence similarity measure and sentence based extractive technique for automatic text summarization. Expert Syst Appl 36(4):7764–7772

Alguliev RM, Aliguliyev RM, Isazade NR (2013) Multiple documents summarization based on evolutionary optimization algorithm. Expert Syst Appl 40:1675–1689. doi:10.1016/j.eswa.2012.09.014

Alguliev RM, Aliguliyev RM, Hajirahimova MS, Mehdiyev CA (2011) MCMR: maximum coverage and minimum redundant text summarization model. Expert Syst Appl 38:14514–14522. doi:10.1016/j.eswa.2011.05.033

Almeida M, Martins AF (2013) Fast and robust compressive summarization with dual decomposition and multi-task learning. In: ACL (1), pp 196–206

Amigó E, Gonzalo J, Penas A, Verdejo F (2005) QARLA: a framework for the evaluation of text summarization systems. In: ACL '05: proceedings of the 43rd annual meeting on association for computational linguistics, pp 280–289

Amati G (2003) Probability models for information retrieval based on divergence from randomness. University of Glasgow

Amini MR, Usunier N (2009) Incorporating prior knowledge into a transductive ranking algorithm for multi-document summarization. In: Proceedings of the 32nd annual ACM SIGIR conference on research and development in information retrieval (SIGIR'09), pp 704–705

Antiqueira L, Oliveira ON, Costa F, Volpe G (2009) A complex network approach to text summarization. Inf Sci 179:584–599. doi:10.1016/j.ins.2008.10.032

Azmi AM, Al-Thanyyan S (2012) A text summarizer for Arabic. Comput Speech Lang 26:260–273. doi:10.1016/j.csl.2012.01.002

Bairi RB, Iyer R, Ramakrishnan G, Bilmes J (2015) Summarization of multi-document topic hierarchies using submodular. In: Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing, pp 553–563

Banerjee S Mitra P, Sugiyama K (2015) Multi-document abstractive summarization using ILP based multi-sentence compression. In: Proceedings of the 24th international joint conference on artificial intelligence (IJCAI 2015), pp 1208–1214

Baralis E, Cagliero L, Jabeen S, Fiori A (2012) Multi-document summarization exploiting frequent itemsets. In: Symposium on applied computing (SAC'12), pp 782–786

Baralis E, Cagliero L, Mahoto N, Fiori A (2013) GRAPHSUM : discovering correlations among multiple terms for graph-based summarization. Inf Sci 249:96–109. doi:10.1016/j.ins.2013.06.046

Barrera A, Verma R (2012) Combining syntax and semantics for automatic extractive single-document summarization. In: 13th international conference on computational linguistics and intelligent text processing. Springer, pp 366–377

Barzilay R, Lapata M (2005) Modeling local coherance: an entity-based approach. In: Proceedings of the 43rd annual meeting of the association for computational linguistics (ACL '05), pp 141–148

Bing L, Li P, Liao Y, Lam W, Guo W, Passonneau RJ (2015) Abstractive multi-document summarization via phrase selection and. arXiv preprint arXiv:1506.01597

Boudin F, Morin E (2013) Keyphrase extraction for N-best reranking in multi-sentence compression. In: North American Chapter of the Association for Computational Linguistics (NAACL)

Brin S, Page L (1998) The anatomy of a large scale hypertextual web search engine. In: Proceedings of the 7th international conference on world wide web 7, pp 107–117

Cao Z, Wei F, Dong L, Li S, Zhou M (2015a) February. Ranking with recursive neural networks and its application to multi-document summarization. In: Twenty-ninth AAAI conference on artificial intelligence

Cao Z, Wei F, Dong L, Li S, Zhou M (2015b) Ranking with recursive neural networks and its application to multi-document summarization. In Twenty-ninth AAAI conference on artificial intelligence

Cao Z, Wei F, Li S, Li W, Zhou M, Wang H (2015c) Learning summary prior representation for extractive summarization. In: Proceedings of ACL: short papers, pp 829–833

Carbonell JG, Goldstein J (1998) The use of MMR, diversity-based re-ranking for re-ordering documents and producing summaries. In: Proceedings of the 21st annual international ACM SIGIR conference on research and development in information retrieval, pp 335–336

Carenini G, Ng RT, Zhou X (2007) Summarizing email conversations with clue words. In: Proceedings of the 16th international conference on World Wide Web. ACM. pp 91–100

Carenini G, Ng RT, Zhou X (2008) Summarizing emails with conversational cohesion and subjectivity. ACL 8:353–361

Carlson L, Marcu D, Okurowski ME (2003) Building a discourse-tagged corpus in the framework of rhetorical structure theory. Springer, Netherlands, pp 85–112

Chali Y, Hasan SA (2012) Query focused multi-document summarization: automatic data annotations and supervised learning approaches. Nat Lang Eng 18:109–145

Chan SWK (2006) Beyond keyword and cue-phrase matching: a sentence-based abstraction technique for information extraction. Decis Support Syst 42:759–777. doi:10.1016/j.dss.2004.11.017

Cilibrasi RL, Vitanyi PMB (2007) The Google similarity distance. IEEE Trans Knowl Data Eng 19:370–383

Deerwester S, Dumais ST, Furnas GW et al (1990) Indexing by latent semantic analysis. J Am Soc Inf Sci Technol 41:391–407

Dunlavy DM, O'Leary DP, Conroy JM, Schlesinger JD (2007) A system for querying, clustering and summarizing documents. Inf Process Manag 43:1588–1605

Erkan G, Radev D (2004) LexRank: graph-based lexical centrality as salience in text summarization. J Artif Intell Res 22:457–479

Fang H, Lu W, Wu F et al (2015) Topic aspect-oriented summarization via group selection. Neurocomputing 149:1613–1619. doi:10.1016/j.neucom.2014.08.031

Fattah MA (2014) A hybrid machine learning model for multi-document summarization. 592–600. doi:10.1007/s10489-013-0490-0

Fattah MA, Ren F (2009) GA, MR, FFNN, PNN and GMM based models for automatic text summarization. Comput Speech Lang 23:126–144. doi:10.1016/j.csl.2008.04.002

Ferreira R, De Souza L, Dueire R et al (2013) Assessing sentence scoring techniques for extractive text summarization. Expert Syst Appl 40:5755–5764. doi:10.1016/j.eswa.2013.04.023

Ferreira R, de Souza Cabral L, Freitas F et al (2014) A multi-document summarization system based on statistics and linguistic treatment. Expert Syst Appl 41:5780–5787. doi:10.1016/j.eswa.2014.03.023

Filippova K (2010) August. Multi-sentence compression: finding shortest paths in word graphs. In: Proceedings of the 23rd international conference on computational linguistics. Association for computational linguistics, pp 322–330

Frank JR, Kleiman-Weiner M, Roberts DA, Niu F, Zhang C, Ré C, Soboroff I (2012) Building an entity-centric stream filtering test collection for TREC 2012. MASSACHUSETTS INST OF TECH CAMBRIDGE

Frey BJ, Dueck D (2007) Clustering by passing messages between data points. Science 315(5814):972–976

Fung P, Ngai G (2006) One story, one flow: hidden Markov Story Models for multilingual multidocument summarization. ACM Trans Speech Lang 3:1–16. doi:10.1145/1149290.1151099

Ganesan K, Zhai C, Han J (2010) Opinosis : a graph-based approach to abstractive summarization of highly redundant opinions. In: Proceedings of the 23rd international conference on computational linguistics, pp 340–348

Genest PE, Lapalme G (2011) Framework for abstractive summarization using text-to-text generation. In: Proceedings of the workshop on monolingual text-to-text generation, Association for Computational Linguistics, pp 64–73

Giannakopoulos G, Karkaletsis V, Vouros G, Stamatopoulos P (2008) Summarization system evaluation revisited: N-gram graphs. ACM Trans Speech Lang Process 5:1–39

Gillick D, Favre B, Hakkani-Tur D, Bohnet B, Liu Y, Xie S (2009) The icsi/utd summarization system at tac 2009. In Proceedings of the text analysis conference workshop, Gaithersburg, MD (USA)

Glavaš G, Šnajder J (2014) Event graphs for information retrieval and multi-document summarization. Expert Syst Appl 41:6904–6916. doi:10.1016/j.eswa.2014.04.004

Goldstein J, Mittal V, Carbonelll J, Kantrowitz M (2000) Multi-document summarization by sentence extraction. In: NAACL-ANLP 2000 workshop on automatic summarization. pp 40–48

Gong Y, Liu X (2001) Generic text summarization using relevance measure and latent semantic analysis. In: Proceedings of the 24st annual international ACM SIGIR conference on research and development in information retrieval. pp 19–25

Graff D, Kong J, Chen K, Maeda K (2003) English gigaword. Linguistic Data Consortium, Philadelphia

Graham Y (2015) Re-evaluating automatic summarization with BLEU and 192 shades of ROUGE. In: Proceedings of the 2015 conference on empirical methods in natural language processing. pp 128–137

Grosz BJ, Weinstein S, Joshi AK (1995) Centering: a framework for modeling the local coherence of discourse. Comput Linguist 21:203–225

Gupta V (2013) Hybrid algorithm for multilingual summarization of Hindi and Punjabi documents. In: Mining intelligence and knowledge exploration. Springer International Publishing, pp 717–727

Gupta V, Lehal GS (2010) A survey of text summarization extractive techniques. J Emerg Technol Web Intell 2:258–268. doi:10.4304/jetwi.2.3.258-268

Gupta P, Pendluri VS, Vats I (2011) Summarizing text by ranking texts units according to shallow linguistic features. In: 13th international conference on advanced communication technology. pp 1620–1625

Haberlandt K, Bingham G (1978) Verbs contribute to the coherence of brief narratives: reading related and unrelated sentence triples. J Verbal Learn Verbal Behav 17:419–425

Hadi Y, Essannouni F, Thami ROH (2006) Unsupervised clustering by k-medoids for video summarization. In: ISCCSP'06 (the second international symposium on communications, control and signal processing)

Halliday MAK, Hasan R (1991) Language, context and text: aspects of language in a social-semiotic perspective. Oxford University Press, Oxford

Harabagiu S, Lacatusu F (2005) Topic themes for multi-document summarization. In: SIGIR' 05: proceedings of the 28th annual international ACM SIGIR conference on research and development in information retrieval. pp 202–209

Harabagiu S, Lacatusu F (2010) Using topic themes for multi-document summarization. ACM Trans Inf Syst 28:13:1–13:47

He T, Shao W, Li F, Yang Z, Ma L (2008) The automated estimation of content-terms for query-focused multi-document summarization. In: Fuzzy systems and knowledge discovery, 2008. FSKD'08. Fifth international conference on IEEE, vol 5, pp 580–584

He Z, Chen C, Bu J, Wang C, Zhang L, Cai D, He X (2012) Document summarization based on data reconstruction. In: AAAI

Hearst M (1997) TextTiling: segmenting text into multi-paragraph subtopic passages. Comput Linguist 23:33–64

Heu JU, Qasim I, Lee DH (2015) FoDoSu: multi-document summarization exploiting semantic analysis based on social Folksonomy. Inf Process Manag 51(1):212–225

Hirao T, Yoshida Y, Nishino M, Yasuda N, Nagata M (2013) Single-document summarization as a tree knapsack problem. EMNLP 13:1515–1520

Hong K, Nenkova A (2014) Improving the estimation of word importance for news multi-document summarization. In: Proceedings of EACL

Hong K, Marcus M, Nenkova A (2015) System combination for multi-document summarization. In: Proceedings of the 2015 conference on empirical methods in natural language processing. pp 107–117

Hovy E, Lin CY, Zhou L, Fukumoto J (2006) Automated summarization evaluation with basic elements. In: Proceedings of the 5th international conference on language resources and evaluation (LREC), pp 81–94

Huang L, He Y, Wei F, Li W (2010) Modeling document summarization as multi-objective optimization. In: Proceedings of the third international symposium on intelligent information technology and security informatics, pp 382–386

Jones KS (2007) Automatic summarising: the state of the art. Inf Process Manag 43:1449–1481. doi:10.1016/j.ipm.2007.03.009

Kabadjov M, Atkinson M, Steinberger J et al. (2010) NewsGist: a multilingual statistical news summarizer. Lecture notes in computer science (including including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics) 6323 LNAI, pp 591–594. doi:10.1007/978-3-642-15939-8_40

Kaljahi R, Foster J, Roturier J (2014) Semantic role labelling with minimal resources: experiments with french. In: Lexical and computational semantics (*SEM 2014), p 87

Kallimani JS, Srinivasa KG, Eswara Reddy B (2011) Information extraction by an abstractive text summarization for an Indian regional language. In: Natural language processing and knowledge engineering (NLP-KE), 2011 7th international conference on IEEE, pp 319–322

Kedzie C, McKeown K, Diaz F (2015) Predicting salient updates for disaster summarization. In: Proceedings of the 53rd annual meeting of the ACL and the 7th international conference on natural language processing. pp 1608–1617

Khan A, Salim N, Jaya Kumar Y (2015) A framework for multi-document abstractive summarization based on semantic role labelling. Appl Soft Comput 30:737–747. doi:10.1016/j.asoc.2015.01.070

Kikuchi Y, Hirao T, Takamura H, Okumura M, Nagata M (2014) Single document summarization based on nested tree structure. In: Proceedings of the 52nd annual meeting of the association for computational linguistics, vol 2, pp 315–320

Kim SM, Hovy E (2005) Automatic detection of opinion bearing words and sentences. In: Companion volume to the proceedings of the international joint conference on natural language processing (IJCNLP), pp 61–66

Kintsch W, Van Dijk TA (1978) Toward a model of text comprehension and production. Psychol Rev 85(5):363

Knuth DE (1977) A generalization of Dijkstra's algorithm. Inf Process Lett 6:1–5

Ko Y, Seo J (2004) Learning with unlabeled data for text categorization using a bootstrapping and a feature projection technique. In: Proceedings of the 42nd annual meeting of the association for computational linguistics (ACL 2004). pp 255–262

Ko Y, Seo J (2008) An effective sentence-extraction technique using contextual information and statistical approaches for text summarization. Pattern Recognit Lett 29:1366–1371. doi:10.1016/j.patrec.2008.02.008

Ko Y, Kim K, Seo J (2003) Topic keyword identification for text summarization using lexical clustering. IEICE Trans Inf Syst E86-D:1695–1701

Kruengkrai C, Jaruskulchai C (2003) Generic text summarization using local and global properties of sentences. In: Proceedings of the ieee/wic international conference on web intelligence (ieee/wic'03)

Kulesza A, Taskar B (2012) Determinantal point processes for machine learning. arXiv preprint arXiv:1207.6083

Kulkarni UV, Prasad RS (2010) Implementation and evaluation of evolutionary connectionist approaches to automated text summarization. J Comput Sci 6:1366–1376

Landauer TK, Foltz PW, Laham D (1998) An intoduction to latent semantic analysis. Discourse Process 25:259–284

Lee DD, Seung HS (1999) Learning the parts of objects by non-negative matrix factorization. Nature 401(6755):788–791

Lee J-H, Park S, Ahn C-M, Kim D (2009) Automatic generic document summarization based on non-negative matrix factorization. Inf Process Manag 45:20–34

Leite DS, Rino LHM (2006) Selecting a feature set to summarize texts in Brazilian Portuguese. Advances in artificial intelligence-IBERAMIA-SBIA 2006:462–471

Li JW, Ng KW, Liu Y, Ong KL (2007) Enhancing the effectiveness of clustering with spectra analysis. IEEE Trans Knowl Data Eng 19:887–902

Li C, Liu F, Weng F, Liu Y (2013) Document summarization via guided sentence compression. In: EMNLP, pp 490–500

Li C, Liu Y, Zhao L (2015a) Using external resources and joint learning for bigram weighting in ilp-based multi-document summarization. In: Proceedings of NAACL-HLT, pp 778–787

Li P, Bing L, Lam W, Li H, Liao Y (2015b) Reader-aware multi-document summarization via sparse coding. arXiv preprint arXiv:1504.07324

Lin CY (2004) ROUGE: a package for automatic evaluation of summaries. In: Proceedings of ACL text summarization workshop, pp 74–81

Lin H, Bilmes J (2010) Multi-document summarization via budgeted maximization of submodular functions. In: Human language technologies: the 2010 annual conference of the North American chapter of the association for computational linguistics, Association for Computational Linguistics, pp 912–920

Lin CY, Hovy E (2000) The automated acquisition of topic signatures for text summarization. In: Proceedings of the 18th conference on computational linguistics, pp 495–501

Liu Y, Wang X, Zhang J, Xu H (2008) Personalized PageRank based multi-document summarization. In: Semantic computing and systems, 2008. WSCS'08. IEEE international workshop on IEEE, pp 169–173

Liu X, Webster JJ, Kit C (2009) An extractive text summarizer based on significant words. In: Proceedings of the 22nd international conference on computer processing of oriental languages, language technology for the knowledge-based economy, Springer, pp 168–178

Liu H, Yu H, Deng ZH (2015) Multi-document summarization based on two-level sparse representation model. In: Twenty-ninth AAAI conference on artificial intelligence

Lloret E, Palomar M (2009) A gradual combination of features for building automatic summarisation systems. Text, speech and dialogue. Springer, Berlin, pp 16–23

Lloret E, Palomar M (2011a) Analyzing the use of word graphs for abstractive text summarization. In: IMMM 2011, first international conference, pp 61–66

Lloret E, Palomar M (2011b) Text summarisation in progress: a literature review. Artif Intell Rev 37:1–41. doi:10.1007/s10462-011-9216-z

Lloret E, Palomar M (2013) Tackling redundancy in text summarization through different levels of language analysis. Comput Stand Interfaces 35:507–518. doi:10.1016/j.csi.2012.08.001

Lloret E, Romá-Ferri MT, Palomar M (2013) COMPENDIUM: a text summarization system for generating abstracts of research papers. Data Knowl Eng 88:164–175. doi:10.1016/j.datak.2013.08.005

Luhn H (1958) The automatic creation of literature abstracts. IBM J Res Dev 2:159–165

Mani I, Maybury M (1999) Advances in automatic text summarization. MIT Press, Cambridge

Manning CD, Raghavan P, Schtze H (2008) Introduction to information retrieval. Cambridge University Press, Cambridge

Mann W, Thompson S (1988) Rhetorical structure theory: toward a functional theory of text organization. Text 8:243–281

Mendoza M, Bonilla S, Noguera C et al (2014) Extractive single-document summarization based on genetic operators and guided local search. Expert Syst Appl 41:4158–4169. doi:10.1016/j.eswa.2013.12.042

Mihalcea R, Tarau P (2004) TextRank: bringing order into texts. In: Conference on empirical methods in natural language processing. pp 404–411

Moawad IF, Aref M (2012) Semantic graph reduction approach for abstractive Text Summarization. In: Proceedings of ICCES 2012, 2012 International Conference on Computer Engineering and Systems, pp 132–138. doi:10.1109/ICCES.2012.6408498

Murdock VG (2006) Aspects of sentence retrieval. University of Massachusetts, Amherst

Neto JL, Freitas AA, Kaestner CAA (2002) Automatic text summarization using a machine learning approach. In: Proceedings of the 16th brazilian symposium on artificial intelligence (sbia), 2507 of lnai. pp 205–215

Neto JL, Santos AD, Kaestner CAA, Freitas AA (2000) Document clustering and text summarization. In: Proceedings of the fourth international conference practical applications of knowledge discovery and data mining (padd-2000), pp 41–55

Nobata C, Satoshi S, Murata M, Uchimoto K, Utimaya M, Isahara H (2001) Sentence extraction system asssembling multiple evidence. In: Proceedings 2nd NTCIR workshop, pp 319–324

Orasan C (2009) Comparative evaluation of term-weighing methods for automatic summarization. J Quant Linguist 16:67–95

Otterbacher J, Erkan G, Radev DR (2009) Biased LexRank: passage retrieval using random walks with question-based priors. Inf Process Manag 45(1):42–54

Oufaida H, Philippe B, Omar Nouali (2015) Using distributed word representations and mRMR discriminant analysis for multilingual text summarization. In: Natural language processing and information systems. Springer International Publishing, pp 51–63

Ouyang Y, Li W, Li S, Lu Q (2011) Applying regression models to query-focused multi-document summarization. Inf Process Manag 47:227–237

Ouyang Y, Li W, Zhang R et al (2013) A progressive sentence selection strategy for document summarization. Inf Process Manag 49:213–221. doi:10.1016/j.ipm.2012.05.002

Owczarzak K (2009) DEPEVAL(summ): dependency-based evaluation for automatic summaries. In: Proceedings of the joint conference of the 47th annual meeting of the ACL and the 4th international joint conference on natural language processing of the AFNLP. pp 190–198

Pang B, Lee L (2008) Opinion mining and sentiment analysis. Found Trends Inf Retr 2:1–135

Pardo TAS, Rino LHM, Nunes MGV (2003a) Neuralsumm: a connexionist approach to automatic text summarization. In: Proceedings of the fourth Brazilian meeting artificial intelligence (ENIA). pp 1–10

Pardo TAS, Rino LHM, Nunes MGV (2003b) Gistsumm: a summarization tool based on a new extractive method. In: Proceedings of the sixth workshop on computational processing of written and spoken portuguese (propor), 2721 of LNAI, pp 210–218

Parveen D, Strube M (2015) Integrating importance, non-redundancy and coherence in graph-based extractive summarization. In: Proceedings of the 24th international conference on artificial intelligence. AAAI Press. pp 1298–1304

Patel A, Siddiqui T, Tiwary US (2007) A language independent approach to multilingual text summarization. In: Large scale semantic access to content (text, image, video, and sound), pp 123–132

Pitler E, Nenkova A (2008) Revisiting readability. In: Proceedings of the 2008 conference on empirical methods in natural language processing. pp 186–195

Prasad RS, Uplavikar NM, Wakhare SS, Jain VY, Avinash T (2012) Feature based text summarization. In: International journal of advances in computing and information researches

Quirk R, Greenbaum S, Leech G (1985) A comprehensive grammar of the English language. Longman, London and New York

Radev D, Tam D (2003) Summarization evaluation using relative utility. In: CIKM '03: proceedings of the 12th international conference on information and knowledge management, pp 508–511

Radev DR, Fan W, Zhang Z, Arbor A (2001) WebInEssence: a personalized web-based multi-document summarization and recommendation system. In: NAACL 2001 workshop on automatic summarization, pp 79–88

Radev D, Allison T, Goldensohn B et al. (2004a) MEAD: a platform for multidocument multilingual text summarization. Proc Lr, 1–4

Radev DR, Jing HY, Stys M, Tam D (2004b) Centroid-based summarization of multiple documents. Inf Process Manag 40:919–938

Riedhammer K, Favre B, Hakkani-Tur D (2010) Long story short- global unsupervised models for keyphrase based meeting summarization. Speech Commun 52:801–815

Rino LHM, Modolo M (2004) Supor: an environment for as of texts in brazilianportuguese. In: Espana for natural language processsing (EsTAL). pp 419–430

Rotem N (2011) Open text summarizer (ots). Retrieved from http://libots.sourceforge.net/

Rush AM, Chopra S, Weston J (2015) A neural attention model for abstractive sentence summarization. arXiv preprint arXiv:1509.00685

Russell SJ, Norvig P (1995) Artificial intelligence: a modern approach. Prentice-Hall International Incorporated, Englewood Cliffs

Sanderson M, Croft WB (1999) Deriving concept hierarchies from text. Proceedings of SIGIR 1999:206–213

Sarkar K (2010) Syntactic trimming of extracted sentences for improving extractive multi-document summarization. J Comput 2:177–184

Shen C, Li T, Ding CH (2011) Integrating clustering and multi-document summarization by bi-mixture probabilistic latent semantic analysis (PLSA) with sentence bases. In: AAAI

Shen D, Sun J-T, Li H et al. (2007) Document summarization using conditional random fields. In: Proceedings of 20th international joint conference on artificial intelligence. pp 2862–2867

Simon I, Snavely N, Seitz SM (2007) Scene summarization for online image collections. In: Computer vision, 2007. ICCV 2007. IEEE 11th international conference on. IEEE. pp 1–8

Sipos R, Shivaswamy P, Joachims T (2012) Large-margin learning of submodular summarization models. In: Proceedings of the 13th conference of the European chapter of the association for computational linguistics, Association for Computational Linguistics, pp 224–233

Song W, Choi LC, Park SC, Ding XF (2011) Fuzzy evolutionary optimization modeling and its applications to unsupervised categorization and extractive summarization. Expert Syst Appl 38:9112–9121

Storn R, Price K (1997) Differential evolution-a simple and efficient heuristic for global optimization over continuous spaces. J Glob Optim 11(4):341–359

Svore K, Vanderwende L, Burges C (2007) Enhancing single-document summarization by combining RankNet and third priority sources. In: Proceedings of the empirical methods on natural language processing and computational natural language learning (EMNLP-CoNLL), pp 448–457

Takamura H, Okumura M (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, Association for Computational Linguistics, pp 781–789

Tan PN, Kumar V, Srivastava J (2002) Selecting the right interestingness measure for association patterns. In: ACM SIGKDD international conference on knowledge discovery and data mining (KDD'02). pp 32–41

Tang J, Yao L, Chen D (2009) Multi-topic based query-oriented summarization. SDM 9:1147–1158

Tao Y, Zhou S, Lam W, Guan J (2008) Towards more text summarization based on textual association networks. In: Proceedings of the 2008 fourth international conference on semantics, knowledge and grid, pp 235–240

Teufel S, Halteren H (2004) Evaluating information content by factoid analysis: human annotation and stability. In: Proceedings of the 2004 conference on empirical methods in natural language processing, pp 419–426

Texlexan (2011) Texlexan: an open-source text summarizer. http://texlexan.sourceforge.net/

Tonelli S, Pianta E (2011) Matching documents and summaries using key concepts. In: Proceedings of the French text mining evaluation workshop

Tzouridis E, Nasir JA, Lahore LUMS, Brefeld U (2014) Learning to summarise related sentences. In: The 25th international conference on computational linguistics (COLING'14), Dublin, Ireland, ACL

Vadlapudi R, Katragadda R (2010) An automated evaluation of readability of summaries: capturing grammaticality, focus, structure and coherence. In: Proceedings of the NAACL HLT 2010 student research workshop. pp 7–12

van der Plas L, Henderson J, Merlo P (2010) D6. 2: semantic role annotation of a French-English Corpus, Computational Learning in Adaptive Systems for Spoken Conversation (CLASSiC)

Van der Plas L, Merlo P, Henderson J (2011) Scaling up automatic cross-lingual semantic role annotation. In: Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies: short papers, vol 2. Association for computational linguistics, pp 299–304

Wan X (2008) Using only cross-document relationships for both generic and topic-focused multi-document summarizations. Inf Retr 11(1):25–49

Wan X (2010) Towards a unified approach to simultaneous single-document and multi-document summarizations. In: Proceedings of the 23rd international conference on computational linguistics (Coling 2010), pp 1137–1145

Wan X, Yang J (2008) Multi-document summarization using cluster-based link analysis. In: Proceedings of the 31st annual international ACM SIGIR conference on research and development in information retrieval. ACM. pp 299–306

Wan X, Xiao J (2009) Graph-based multi-modality learning for topic-focused multi-document summarization. In: IJCAI. pp. 1586–1591

Wang D, Li T (2012) Weighted consensus multi-document summarization. Inf Process Manag 48:513–523

Wang C, Long L, Li L (2008a) HowNet based evaluation for Chinese text summarization. In: Proceedings of the international conference on natural language processing and software engineering. pp 82–87

Wang D, Li T, Zhu S, Ding C (2008b) Multi-document summarization via sentence-level semantic analysis and symmetric matrix factorization. In: Proceedings of the 31st annual international ACM SIGIR conference on research and development in information retrieval, pp 307–314

Wang D, Li T, Zhu S, Ding C (2009) Multi-document summarization using sentence-based topic models. In: Proceedings of the ACL-IJCNLP 2009 conference short papers, pp 297–300

Wang D, Li T, Ding C (2010) Weighted feature subset non-negative matrix factorization and its applications to document understanding. In: Proceedings of the 2010 IEEE international conference on data mining, pp 541–550

Wang D, Zhu S, Li T et al. (2011) Integrating document clustering and multi-document summarization. ACM Trans Knowl Discov Data 5:14:1–14:26

Wasson M (1998) Using leading text for news summaries: evaluation results and implications for commercial summarization applications. In: Proceedings of the 17th international conference on computational linguistics, vol 2. Association for computational linguistics, pp 1364–1368

Wei F, Li W, Lu Q, He Y (2008) Query sensitive mutual reinforcement chain and its application in query-oriented multi-document summarization. In: Proceedings of the 31st annual international acmsigir conference on research and development in information retrieval (SIGIR'08). pp 283–290

Wei F, Li W, Lu Q, He Y (2010) A document-sensitive graph model for multi-document summarization. Knowl Inf Syst 22(2):245–259

Wenjie L, Furu W, Qin L, Yanxiang H (2008) Pnr2: ranking sentences with positive and negative reinforcement for query-oriented update summarization. In: Proceedings of the 22nd international conference on computational linguistics (coling'08). pp 489–496

Wilson T, Hoffmann P, Somasundaran S, Kessler J, Wiebe J, Choi Y, Cardie C, Riloff E, Patwardhan S (2005) OpinionFinder: a system for subjectivity analysis. In: Proceedings of hlt/emnlp on interactive demonstrations. Association for computational linguistics. pp 34–35

Yang CC, Wang FL (2008) Hierarchical summaization of large documents. J Am Soc Inf Sci Technol 59:887–902

Yang C, Shen J, Peng J, Fan J (2013) Image collection summarization via dictionary learning for sparse representation. Pattern Recognit 46(3):948–961

Yang L, Cai X, Zhang Y, Shi P (2014) Enhancing sentence-level clustering with ranking-based clustering framework for theme-based summarization. Inf Sci 260:37–50. doi:10.1016/j.ins.2013.11.026

Yao JG, Wan X, Xiao J (2015a) Compressive document summarization via sparse optimization. In: Proceedings of the 24th international conference on artificial intelligence. AAAI Press. pp 1376–1382

Yao JG, Wan X, Xiao J (2015b) Phrase-based compressive cross-language summarization. In: Proceedings of the 2015 conference on empirical methods in natural language processing, pp 118–127

Ye S, Chua TS, Kan MY, Qiu L (2007) Document concept lattice for text understanding and summarization. Inf Process Manag 43:1643–1662. doi:10.1016/j.ipm.2007.03.010

Yeh J-Y, Ke H-R, Yang W-P, Meng I-H (2005) Text summarization using a trainable summarizer and latent semantic analysis. Inf Process Manag 41:75–95. doi:10.1016/j.ipm.2004.04.003

Yen JY (1971) Finding the k shortest loopless paths in a network. Manag Sci 17(11):712–716

Zajic DM, Dorr BJ, Lin J (2008) Single-document and multi-document summarization techniques for e-mail threads using sentence compression. Inf Process Manag 44:1600–1610

Zha H (2002) Generic summarization and keyphrase extraction using mutual reinforcement principle and sentence clustering. In: Proceedings of the 25th annual international acmsigir conference on research and development in information retrieval (SIGIR'02), pp 113–120

Zhang J, Xu H, Cheng X (2008a) Gspsummary: a graph-based sub-topic partition algorithm for summarization. In: Proceedings of the 2008 Asia information retrieval symposium, pp 321–334

Zhang J, Cheng X, Wu G, Xu H (2008b) Ada sum: an adaptive model for summarization. In: Proceedings of the acm 17th conference on information and knowledge management (CIKM'08), pp 901–909

Zhao L, Wu L, Huang X (2009) Using query expansion in graph-based approach for query-focused multi-document summarization. Inf Process Manag 45(1):35–41

Zhou L, Lin CY, Munteanu DS, Hovy E (2006) ParaEval: using paraphrases to evaluate summaries to evaluate summaries automatically. In: Proceedings of the human language technology/North American association of computational linguistics conference, pp 447–454