

Unsupervised Broadcast News Summarization; a comparative study on Maximal Marginal Relevance (MMR) and Latent Semantic Analysis (LSA)

Majid Ramezani

Computerized Intelligence Systems
Laboratory

Department of Computer Engineering
University of Tabriz
Tabriz, Iran

m_ramezani@tabrizu.ac.ir

Mohammad-Salar Shahryari

Department of Computer Engineering
University of Tabriz

Tabriz, Iran

s.shahryari@tabrizu.ac.ir

Amir-Reza Feizi-Derakhshi

Computerized Intelligence Systems
Laboratory

Department of Computer Engineering
University of Tabriz
Tabriz, Iran

derakhshi97@ms.tabrizu.ac.ir

Mohammad-Reza Feizi-Derakhshi

Computerized Intelligence Systems
Laboratory

Department of Computer Engineering
University of Tabriz

Tabriz, Iran

mfeizi@tabrizu.ac.ir

Abstract— The automatic speech summarization methods traditionally are classified into two groups: supervised and unsupervised methods. Supervised methods rely on a set of features, while unsupervised methods perform summarization through a set of rules. Among unsupervised automatic speech summarization methods, Latent Semantic Analysis (LSA) and Maximal Marginal Relevance (MMR) are so famous. This study set out to peruse the overall efficacy of two aforementioned unsupervised methods in summarization of Persian broadcast news transcriptions. The results justify the superiority of LSA to MMR during generic summarization. This is while MMR achieves better results in query-based summarization.

Keywords—broadcast news summarization, unsupervised summarization, Maximal Marginal Relevance (MMR), Latent Semantic Analysis (LSA)

I. INTRODUCTION

Undoubtedly, the role of information cannot be ignored in today's human life. Concerning the fact that the available information is more than what is required, and the volume of information increases exponentially, there should be some tools by which the essential information can be acquired in various information resources. In this regard, automatic summarization systems aim to extract the most essential and available information from one or more resources. They are considered a competent substitution for perusing the whole document. In addition, they cause to save a lot of time.

Research on automatic summarization began in 1950 in the field of automatic summarization of text. It has been considerably taken into account in recent years ([1]-[11]). Successes in automatic text summarization encouraged the researchers toward automatic speech summarization which refers to extracting the most critical information available in speech documents and removing redundant and irrelevant information [12]. Therefore, gaining access to the required information and searching information among great speech documents such as broadcast news, lectures, presentations, multi-party meeting recordings, spontaneous dialogues, voicemail, telephone conversations etc. have been facilitated [13]-[17].

Generally, speech summarization methods are classified into two groups. Namely, *supervised* (or *feature-based*) methods such as sentence position techniques, cue words, and acoustic/prosodic features like duration and pitch, and *unsupervised* (or *rule-based*) methods such as *Maximal Marginal Relevance (MMR)*, and *Latent Semantic Analysis (LSA)* [18].

A considerable amount of literature has been published on the speech summarization of broadcast news. In these studies, text transcripts of speech documents are available. For this purpose, both text-based and speech-based features have been utilized. Some researchers have investigated and studied speech summarization by using supervised methods. A set of linguistic features (such as word frequency, cue words, title words, and location) and non-linguistic features (such as prosodic information) have been used in [19] to summarize Japanese broadcast news. Banerjee and Rudnicky [20] have developed an extractive summarization system in multi-party human-human dialogue using text-based features. Various information, containing lexical and structural information and acoustic/prosodic information, have been used by [21] to summarize broadcast news. Other researchs, such as [22] have investigated different features like discourse, structural and lexical features, and acoustic/prosodic features to summarize English and Japanese broadcast news. By using a Support Vector Machine (SVM), authors in [23] proposed a method to summarize speech documents. The utterance importance and redundancy have also been simultaneously taken into account. Liu and Liu [24] developed a system by applying a rich set of speaker-sensitive features to summarize meeting speech documents. Besides, Wang and Cardie [25] proposed an abstract generation framework regardless of the domains of the inputs. Zhang et al. in [26] leveraged rhetorical structures of speech documents to summarize lecture speeches.

In addition, various studies have been conducted using unsupervised methods of speech summarization. ClusterRank [27] is a graph-based system that is proposed for extractive meeting summarization; it uses an extension of the TextRank [28] algorithm (an algorithm for sentence extraction from the text). TextRank uses Google's PageRank to calculate a

(relevance) score for each sentence. Authors in [29] considered a graph-based method to perform speech summarization. They combined prosodic features with an unsupervised speech summarization framework. Chen et al. [30] and Lee et al. [31] used graph-based methods in lecture summarization. Other unsupervised methods, like Language Modeling Approach (LM) and Topical Mixture Model (TMM) are used in automatic summarization to predict terms using n-grams [32]. Among unsupervised methods, LSA and MMR are regarded as the most well-known and widely-used methods in transcript summarization. Extractive summarization follows two objectives: maximizing the amount of covered information and minimizing redundancy [33]. Basically, MMR [34] is a greedy and query-based method that performs summarization through sentence-to-sentence selection, bearing in mind two essential criteria, namely similarity to the user query and minimum redundancy in selected sentences. Applying the same strategy the authors in [35] leveraged several steps including speech disfluency and removal, sentence boundary detection, identification, linking query-answer regions, and topic segmentation to summarize the spoken dialogues. The authors in [36] by identifying and extracting key phrases, focused on finding the best queries to apply in MMR.

LSA [37] leveraging the contextual usage of words, determines the implicit semantic relations among them. The authors in [38] to perform single and multi-documents of broadcast news summarization, applied LSA to identify and extract the most prominent topics and concepts that appeared in the input documents. Their proposed system performs summarization without employing a lexical thesaurus, such as WordNet. In the same manner the authors in [39] leveraged LSA in terms of story segmentation of Chinese broadcast news. In another attempt the authors in [40] proposed a system to reduce the negative effects of synonyms in information retrieval and to identify the story boundaries in Malay broadcast news summarization. Hofmann has proposed a technique called Probabilistic LSA [41]. As can be inferred, it has been derived from the LSA method. This technique involves a stronger statistical basis than LSA.

This study aims to perform unsupervised broadcast news summarization using MMR and LSA methods and evaluate their success. For this purpose, Persian broadcast news is used. In section 2, MMR and LSA are explained, and their theories are studied in detail. Then, experimental results obtained from the implementation of an automatic summarization system in broadcast news are presented in section 3, and the obtained results will be analyzed. Finally, conclusions are presented in section 4.

II. METHODOLOGY

In this study, the summarization of broadcast news will be presented by using MMR and LSA. Generally, summarizations can be performed in both *generic* and *query-based* ways. Generic summarizations include extracting the most important information available in a document. Summaries produced using generic summarization are considered a complete substitution of original documents. In contrast, in query-based summarizations, summaries are presented based on the required information of the user determined by the query. In the following sections, summarizations using LSA or MMR methods are investigated in terms of generic and query-based methods.

A. Maximal Marginal Relevance (MMR)

Maximal Marginal Relevance (MMR) is one of the most well-known methods in information retrieval to balance *relevance* and *diversity* [42]. This method was proposed by Carbonell and Goldstein [34] for the first time. It selects the sentences based on a weighted combination of their relevance to the input query and their similarity to the sentences that have been selected before. The idea behind this method is the iterative ranking of sentences according to the input query. In each iteration, a sentence is selected in which: it has not been selected before, it is more similar to the query, and it is more dissimilar to those previously selected sentences. According to this method, the MMR score is calculated by (1) for each S_i sentence of the input document.

$$Score^{MMR}(S_i) = \operatorname{argmax}_{S_i} \{ \lambda (Sim_1(S_i, Q)) - (1-\lambda) (Sim_2(S_i, Summ)) \} \quad (1)$$

where Q is the query vector, $Summ$ is the vector of those sentences selected to appear in the final summary, and λ is a coefficient between 0 and 1, which controls the amount of redundancy and similarity to the input query (relevance) in each selection. Sim_1 measures the similarity between the user query and current sentences, while Sim_2 evaluates the similarity of the current sentence with those sentences that have been selected up to now. In this method, the cosine similarity function has been used as a similarity metric for Sim_1 and Sim_2 . The sentences with the highest MMR score will be iteratively selected to appear in the summary until it reaches a predefined proper size. In this regard, the λ coefficient has great importance, and the quality of results depends on the value of this coefficient. As it can be inferred, there is a trade-off between similarity to the query (relevance) and diversity. This relation is affected by the value of λ . If $\lambda=1$, then the relevance reaches its maximum value. If $\lambda=0$, then diversity will be maximized.

Despite that, MMR basically is a query-based method, it can be used as a generic method [43]. In such a case, the document vector (D) substitutes the query vector (Q) in (1).

B. Latent Semantic Analysis (LSA)

Latent semantic Analysis (LSA) is a method that is used to extract and represent the contextual usage of words [44]. This way, semantic similarities among the words and other sets of words are identified. LSA is a vector space method that can extract word-word, word-passage, and passage-passage relations. There are high correlations among these relations, human cognitive events, and semantic representation [45]. This method can considerably extract and identify semantic relations formed in the speaker or writer's mind during writing and speaking. Also, it can be mentioned that LSA allows us to provide semantic similarities among words for the machine through proper approximation of human judgment and to predict semantic relations between various text parts.

LSA is a full-automatic mathematical-statistical method composed of two basic steps. The first step involves representing the input text by a matrix whose rows are related to the input text's unique words. The columns are dedicated to sentences, passages, or other text units (such as paragraphs or documents). In the first step, LSA creates a *term-by-sentence* matrix (A), which is a matrix representation of the input text. Each column vector of the matrix (A_i) is an indication of a weighted term-frequency vector for the i^{th} sentence. Having t unique words/terms and s sentences in the input text, A would be a $t \times s$ matrix ($t > s$). Each cell of the matrix involves the

frequency of the words/terms in the corresponding sentence. Of course, there are various approaches for determining matrix cells, such as the number of occurrences, the binary representation of the number of occurrences, TF- IDF (Term Frequency-Inverse Document Frequency), Log Entropy, Root Type, and Modified TF-IDF.

The second step of LSA is applying the Singular Value Decomposition (SVD) operator over matrix A . As a consequence of applying SVD, matrix A is decomposed into three matrices, namely, V , Σ and U .

$$A_{t \times s} = U_{t \times c} \Sigma_{c \times c} V^T_{c \times s} \quad (2)$$

Where U is a $t \times c$ and column-orthogonal matrix whose columns are called *left singular values*. $\Sigma = \text{diag}(\delta_1, \delta_2, \dots, \delta_c)$ is a diagonal and $c \times c$ matrix whose main diagonal elements are *eigenvalues* of A . The eigenvalues have been placed in descending order on the main diagonal. V^T is an orthogonal $c \times s$ matrix whose columns are called *right singular values*. If $\text{rank}(A)=r$, then the following relation is presented for the matrix Σ :

$$\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \dots \geq \sigma_r \geq \sigma_{r+1} = \dots = \sigma_s = 0$$

Matrices obtained from applying the SVD operator to the input matrix can be explained as follows [46]: each column in U , which is a $t \times c$ matrix, indicates a unique concept or topic of the input text. In this matrix, the rows are the words or terms available in the original space. Each cell indicates the weight of the corresponding term in the corresponding concept or topic. The diagonal matrix Σ with $c \times c$ dimensions involves the significance of each concept or topic that appeared in the original text. Since the cells are sorted in descending order, the concepts with less importance can be ignored by reduction. Moreover, V^T is a new representation of sentences in the input text. The difference is that, in this representation, sentences are not represented using words and terms in the original input text, but they are represented using the terms of the topics given in U . Keeping in mind that although that LSA is a generic method, it can also be used as a query-based method. They are described in detail below.

1) Generic LSA

For the first time, Gong and Liu [38] proposed the application of LSA for automatic summarization. Subsequently, Steinberger and Jezek [47] and Murray et al. [48] extended and improved the LSA-based summarization. As explained below, all of the proposed methods apply the two main steps in LSA along with a sentence selection step. Sentence selection is the central focus of all LSA-based methods in the current study.

a) Gong and Liu

The study by Gong and Liu [55] is considered the most crucial in LSA-based automatic summarization. According to their suggestion, the final summary is acquired after the creation of $t \times s$ matrix and decomposing it by applying the SVD. Then the summary will be acquired using a sentence selection step which primarily relies on the V^T matrix. According to their algorithm to select sentences, the first concept (that is, the most important concept, which is placed in the first row of V^T) is selected first. Then, the sentence most relevant to this concept (the sentence with the highest value in the corresponding row) will be presented in the final summary. It continues for other concepts so that a predefined number of sentences are extracted. Fig. 1, demonstrates an example of V^T matrix. According to this algorithm, the *Con. 0*

concept is first selected, and then S_1 (the cell with the highest value in *Con. 0*) is selected for the final summary.

| | S_0 | S_1 | S_2 | S_3 | S_4 |
|---------------|-------|-------|-------|-------|-------|
| Con. 0 | 0.557 | 0.691 | 0.241 | 0.110 | 0.432 |
| Con. 1 | 0.345 | 0.674 | 0.742 | 0.212 | 0.567 |
| Con. 2 | 0.732 | 0.232 | 0.435 | 0.157 | 0.246 |
| Con. 3 | 0.628 | 0.836 | 0.783 | 0.265 | 0.343 |

Fig. 1. An example of V^T matrix (Gong & Liu). From each row of V^T , the sentence with the highest score is selected until a predefined number of sentences are selected.

| | <i>Con. 0</i> | <i>Con. 1</i> | <i>Con. 2</i> | <i>Con. 3</i> | <i>Length</i> |
|-------|---------------|---------------|---------------|---------------|---------------|
| S_0 | 0.846 | 0.334 | 0.231 | 0.210 | 0.432 |
| S_1 | 0.455 | 0.235 | 0.432 | 0.342 | 0.543 |
| S_2 | 0.562 | 0.632 | 0.735 | 0.857 | 0.723 |
| S_3 | 0.378 | 0.186 | 0.248 | 0.545 | 0.235 |

Fig. 2. An example of V matrix (Steinberger & Jezek). For each row of V , the length of the sentences using n concepts is calculated. n is given as an input parameter. The values in Σ are used as importance while achieving the lengths.

| | S_0 | S_1 | S_2 | S_3 | S_4 |
|---------------|-------|-------|-------|-------|-------|
| Con. 0 | 0.557 | 0.691 | 0.241 | 0.110 | 0.432 |
| Con. 1 | 0.345 | 0.674 | 0.742 | 0.212 | 0.567 |
| Con. 2 | 0.732 | 0.232 | 0.435 | 0.157 | 0.246 |
| Con. 3 | 0.628 | 0.836 | 0.783 | 0.265 | 0.343 |

Fig. 3. An example of V^T matrix (Murray et al.). From each row of V^T , one or more sentences with higher scores are selected. The number of sentences to be selected is determined by using Σ .

The algorithm suffers from some defects. Firstly, the size of the reduced dimension should be equal to the final summary's length. Therefore, selecting some sentences with less significance is probable. Secondly, some sentences are relevant to the selected concept, but the related cell does not have the highest value in the corresponding row. According to this algorithm, such sentences cannot be presented in summary (while such sentences result in increasing the relevance in the final summary). Thirdly, unlike the LSA, the significance of all concepts is considered equally, while some of them don't have great importance in the original document.

b) Steinberger and Jezek

Steinberger and Jezek [47] proposed an algorithm to improve the defects of the method suggested by Gong and Liu. In this algorithm, after performing two basic steps of LSA, Σ and V matrices have been used to select the sentences. In this algorithm, the sentence selection begins with calculating the length of sentence vectors placed in the rows of the V . To calculate the length of a sentence vector, the concepts whose index is less than or equal to the number of dimensions in the new space are used. The size of new space dimensions is considered as input in this algorithm. Σ is a diagonal matrix that entails the eigenvalues of A , which are sorted in descending order in the main diagonal. This matrix conceptually involves the amount of importance of the concepts proposed in the original document that has been transferred to the reduced dimension space. Therefore, by considering the values in this matrix as a multiplication parameter, these concepts that are more relevant to the document will have great importance. If the dimension of the new space is equal to n , then the length of the i^{th} sentence is calculated according to (3).

$$\text{Length } i = \sqrt{\sum_{j=1}^n V_{ij} \times \Sigma_{jj}} \quad (3)$$

After calculating the length of all sentences, the longest sentences are selected for the summary. As it can be inferred, the selection of more than one sentence from each concept is possible in this algorithm. Sentences that are more relevant to significant concepts are selected as well. It's evident that in this algorithm, the significance of various concepts extracted by SVD is not considered equal. Fig. 2, shows a sample V matrix. If the dimension of the new space is equal to 3, then the length of sentences is calculated by considering the first 3 concepts, and finally, S_2 is selected as the first sentence.

c) Murray et al.

To solve the drawbacks of the algorithm proposed by Gong and Liu, Murray et al. [48] also attempted to propose an algorithm. Likewise, their algorithm at first performs the two basic steps of LSA and then tries to perform its specific sentence selection strategy. This algorithm is similar to the algorithm proposed by Gong and Liu, but the main difference is that here, n number of the best sentences are selected instead of selecting the best sentence from each concept. The value of n is calculated by singular values in the matrix Σ . In this algorithm, Σ and V^T are used in sentence selection. The number of sentences selected from each concept is determined by calculating the percentage of corresponding singular value in Σ over the sum of all singular values in Σ . Therefore, selecting more than one sentence from each concept is possible. In addition, the significance of various extracted concepts is not considered equal. In addition, dimension reduction is not dependent on the length of the final summary. Fig. 3, shows a sample V^T matrix. Suppose that two sentences must be selected from *Con. 0*, and one sentence must be selected from the *Con. 1*. According to this algorithm, S_0 and S_1 are selected from *Con. 0*, and S_2 is selected from *Con. 1*.

2) Query-based LSA

In the query-based approach, the obtained results are affected by the user query. Generally, the method of literally matching the terms in the user's query and terms in documents can apply user query to the information retrieval. But it should be noted that literal or lexical matching and correspondence involve incorrect results [37]. Usually, there are different methods for stating a concept. Concerning this issue, such documents can not be identified and considered relevant. In addition, most of the words are polysemous. Therefore, it's possible to compare the words of the user's query with the words of irrelevant documents. In the query-based LSA to solve the problems of lexical matching, documents as well as the user's query, are projected to the new space. In this new space, there may be a high cosine similarity between a query and a document without any common word. Because according to LSA, the words of query and sentences may be semantically similar. It can be stated that the LSA method represents documents and queries in reduced dimensional space by using the SVD operator instead of representing them based on vectors with independent words. Since the number of factors or dimensions in the new space is less than when the words are separately represented, the words are not independent. For instance, if two words are presented in a similar context, they will be represented by a single vector in reduced dimensional space. Accordingly, even those sentences having no shared words with the input query could be retrieved effectively. Same to the generic approach, we also need to represent the input text in a matrix for query-based LSA. Therefore the matrix will be projected to a latent semantic space by applying the SVD (refer to (2)).

Finally, the user query q should be transferred to the new space using (4). It also provides a logical and mathematical basis for evaluating the similarity of query vectors (\hat{q}) and sentence vectors in new semantic space.

$$\hat{q} = q^T U_{t \times r} \Sigma_{r \times r}^{-1} \quad (4)$$

where q is the $t \times 1$ vector of the user's query. Then, multiplying $q^T U_{t \times r} \Sigma_{r \times r}^{-1}$, the query vector is transferred from the original space to the new space. Finally, multiplying it by $\Sigma_{r \times r}^{-1}$, suitable weights are assigned to each dimension. In this regard, the required basis for comparing query vectors with sentences' vectors is provided in new space. For this purpose, n number of sentences in the input text whose cosine similarity is higher than the required threshold is selected for the summary.

3) Probabilistic LSA (PLSA)

Probabilistic Latent Semantic Analysis (PLSA) [41] is a method obtained from the LSA technique by emphasizing statistical aspects. In other words, this technique has a stronger statistical basis by emphasizing the Maximum Likelihood Principle. PLSA uses Hofmann's Aspect Model. Based on this method, a set of unobserved variables, $z \in Z = \{z_1, z_2, \dots, z_k\}$ that appeared topics in documents are related to two sets of observed variables including documents $d \in D = \{d_1, d_2, \dots, d_m\}$ that here are considered as sentences, and the words of these documents $w \in W = \{w_1, w_2, \dots, w_n\}$. In this method the probability of each of the co-occurrences is computed using (5) by observing the co-occurrence of the words and documents (w, d):

$$P(d, w) = P(d) P(w|d), \quad P(w|d) = \sum_{z \in Z} P(w|z) P(z|d) \quad (5)$$

where $P(d)$ is the probability of selecting document d , $P(w|z)$ is the probability of selecting the word w from topic z , $P(z|d)$ stands for the probability of selecting topic z in document d , and $P(d, w)$ shows the likelihood of each document and word pair. In the next step, the values of $P(d)$, $P(w|z)$ and $P(z|d)$ are optimized by using the maximization of log-likelihood based on the Maximum Likelihood Principle, and the maximum value is determined. For this purpose, the expectation-maximization algorithm can be used. This algorithm is an iterative method in statistics, and it is used to determine the Maximum Likelihood when the model depends on a set of unobserved latent variables [41]. This algorithm involves two steps: the expectation step (*E-step*), in which the previous probabilities related to latent variables are computed, and the Maximization step (*M-step*), in which parameters are updated. Hofmann has proposed the (6) for E-step:

$$P(z|d, w) = \frac{P(z)P(d|z)P(w|z)}{\sum_{z \in Z} P(z)P(d|z)P(w|z)} \quad (6)$$

In addition, he has suggested the (7), (8) and (9) for M-step:

$$P(w|z) = \frac{\sum_d n(d, w) P(z|d, w)}{\sum_{d, w} n(d, w) P(z|d, w)} \quad (7)$$

$$P(d|z) = \frac{\sum_w n(d, w) P(z|d, w)}{\sum_{d, w} n(d, w) P(z|d, w)} \quad (8)$$

$$P(z) = \frac{\sum_{d, w} n(d, w) P(z|d, w)}{\sum_{d, w} n(d, w)} \quad (9)$$

where, $n(d, w)$ is the frequency of w in d .

E-step includes computing the probability of that word w appeared in document d can be explained by the factor z (topic). Then, the converging point can be found by alternating the E-step and M-step. This point has the role of local maximum in log-likelihood.

The PLSA-based summarization algorithm involves the following steps:

- i. Each document is represented in terms of the term-frequency matrix.
- ii. The values of $P(z)$, $P(d|z)$ and $P(w|z)$ are calculated as declared in (7), (8), and (9) until the convergence criterion for EM-algorithm is obtained. $P(d|z)$ shows the importance of document d (sentence) in the given topic represented by z . $P(z)$ represents the importance of topic z in document d .
- iii. R score is computed by using the (10) for each sentence. In this case, all sentences are scored according to their topic inclusion.

$$R = \sum_z P(d|z)P(z) = P(d) \quad (10)$$

- iv. Finally, the sentences with the highest R score are selected to form the summary.

III. EXPERIMENTS AND RESULTS

A. Dataset

The corpus used in this study results from the human transcription of Persian broadcast news for more than 15 hours, containing 58 news documents. Four native transcribers have collected it from the news of three radio channels and four TV channels over 45 days. The corpus involves more than 115,000 words and 7,000 sentences. The average number of sentences per news document is 122.5, and the average length of sentences in the corpus is 16.5. To produce the golden summaries as the references to evaluate summarization systems, four human summarizers summarized all the documents, for both generic and query-based summarization (for pre-specified queries). The average number of sentences per golden summary is equal to 36.7.

B. Results and Discussions

To evaluate the proposed summarization systems, we used *Precision*, *Recall* and *F1-score*, along with the *ROUGE-1* with a compression rate equal to 30%.

$$Precision = \frac{TP}{TP + FP} \quad (11)$$

$$Recall = \frac{TP}{TP + FN} \quad (12)$$

$$F1\text{-score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (13)$$

$$ROUGE\text{-}n = \frac{\sum_{S \in \{ReferenceSummaries\}} \sum_{gram_n \in S} Count_{match}(gram_n)}{\sum_{S \in \{ReferenceSummaries\}} \sum_{gram_n \in S} Count(gram_n)} \quad (14)$$

where, TP , FP , and FN respectively, denote the number of true positives, false positives, and false negatives. In (14) the size of the n -gram is denoted by n (here $n=1$). $Count_{match}(gram_n)$ stands for the maximum number of n -grams co-occurring in the candidate summary, and $Count(gram_n)$ is the number of n -grams in the reference summary.

1) Summarization Using MMR

As mentioned before, in MMR documents are summarized to provide a balance between relevance and redundancy. It is followed by setting λ (see (1)). The highest value of λ will cause more emphasis on relevance. And the lowest value of λ ,

TABLE I. EVALUATION RESULTS FOR QUERY-BASED MMR

| λ | Precision (%) | Recall (%) | F-score (%) | ROUGE-1 |
|-----------|---------------|--------------|--------------|---------------|
| 0.0 | 33.34 | 21.87 | 26.22 | 0.2437 |
| 0.1 | 35.72 | 23.54 | 28.22 | 0.2494 |
| 0.2 | 35.72 | 23.54 | 28.22 | 0.2267 |
| 0.3 | 65.87 | 46.25 | 54.17 | 0.4310 |
| 0.4 | 71.43 | 49.28 | 58.09 | 0.4487 |
| 0.5 | 61.90 | 42.05 | 49.88 | 0.3982 |
| 0.6 | 76.19 | 49.26 | 59.56 | 0.4847 |
| 0.7 | 72.22 | 48.79 | 57.97 | 0.4839 |
| 0.8 | 69.84 | 46.94 | 55.88 | 0.4681 |
| 0.9 | 69.84 | 46.94 | 55.88 | 0.4681 |
| 1.0 | 83.34 | 54.66 | 65.69 | 0.5379 |

TABLE II. EVALUATION RESULTS FOR GENERIC MMR

| λ | Precision (%) | Recall (%) | F-score (%) | ROUGE-1 |
|-----------|---------------|--------------|--------------|---------------|
| 0.0 | 22.22 | 20.04 | 21.06 | 0.2082 |
| 0.1 | 24.61 | 22.26 | 23.36 | 0.2249 |
| 0.2 | 19.84 | 18.10 | 18.92 | 0.1678 |
| 0.3 | 37.30 | 33.13 | 35.08 | 0.3345 |
| 0.4 | 38.10 | 33.45 | 35.61 | 0.3227 |
| 0.5 | 33.32 | 29.56 | 31.32 | 0.2921 |
| 0.6 | 26.19 | 23.18 | 24.58 | 0.2275 |
| 0.7 | 40.48 | 35.68 | 37.91 | 0.3807 |
| 0.8 | 40.48 | 35.68 | 37.91 | 0.3807 |
| 0.9 | 40.48 | 35.68 | 37.91 | 0.3957 |
| 1.0 | 40.48 | 35.68 | 37.91 | 0.3957 |

will cause more diversity. In this study, we considered the results obtained from this method by assigning different values of $\{0, 0.1, 0.2, \dots, 1.0\}$ to λ . Table I demonstrates evaluation results obtained by using query-based MMR for four evaluation metrics. The applied queries are predefined for each news document in the dataset.

As it can be seen, by increasing λ , all of the evaluation metrics were improved so that the best results were obtained in $\lambda=1$. Therefore, it can be inferred that the best results are obtained in broadcast news summarization by using query-based MMR when just the similarity of selected sentences with the query is considered ($\lambda=1$).

Table II demonstrates the results obtained from the evaluation of summaries produced by using generic MMR for four evaluation metrics. As can be seen, like generic MMR, by increasing λ , the performance of the system improves so that the best results are obtained from the maximum value of λ (0.7, 0.8, 0.9, and 1.0). Comparing the query-based MMR and generic MMR reveals that query-based MMR generally outperforms generic MMR.

2) Summarization Using LSA

Table III shows the evaluation results of broadcast news using generic and query-based LSA. Among generic approaches of LSA, the algorithm proposed by Murray et al. has achieved the best results. As can be seen, generally generic LSA outperforms the query-based LSA in all evaluation metrics. Table IV depicts the evaluation results for PLSA-based summarization. Summarization was done for various values of $n=2,3,4,5,6$. Since in $n=1$, only the sentences of one topic are selected for the final summary, the value of $n=1$ has not been taken into account. As can be observed, increasing the value of n until $n=3$ leads to achieving the best results.

3) MMR vs LSA: Which Is the Best Approach in Broadcast News Summarization?

Table V compares generic LSA and MMR, and Table VI compares query-based LSA and MMR. As it is observed, between generic MMR and query-based MMR, better results are obtained in query-based MMR in terms of broadcast news

TABLE III. EVALUATION RESULTS FOR GENERIC LSA AND QUERY-BASED LSA

| | Method | Precision (%) | Recall (%) | F-score (%) | ROUGE-1 |
|---------|---------------------------|---------------|--------------|--------------|---------------|
| Generic | Gong & Liu | 55.96 | 49.39 | 52.46 | 0.5737 |
| | Steinberger & Jezek (n=1) | 57.14 | 50.68 | 53.70 | 0.5339 |
| | Murray et al. | 60.32 | 53.35 | 56.61 | 0.5930 |
| | Query-Based | 54.76 | 42.59 | 47.08 | 0.4854 |

TABLE IV. EVALUATION RESULTS FOR PLSA

| n | Precision (%) | Recall (%) | F-score (%) | ROUGE-1 |
|---|---------------|--------------|--------------|---------------|
| 2 | 58.31 | 28.96 | 38.39 | 0.3167 |
| 3 | 43.32 | 38.08 | 40.48 | 0.4012 |
| 4 | 53.33 | 23.97 | 32.57 | 0.2507 |
| 5 | 50.83 | 21.47 | 29.56 | 0.2258 |
| 6 | 48.32 | 19.48 | 38.68 | 0.2035 |

TABLE V. COMPARING GENERIC LSA WITH GENERIC MMR

| Method | Precision (%) | Recall (%) | F-score (%) | ROUGE-1 |
|-----------------------|---------------|--------------|--------------|---------------|
| LSA (Murray et al.) | 60.32 | 53.35 | 56.61 | 0.5930 |
| MMR ($\lambda=1.0$) | 40.48 | 35.68 | 37.91 | 0.3957 |

TABLE VI. COMPARING QUERY-BASED LSA WITH QUERY-BASED MMR

| Method | Precision (%) | Recall (%) | F-score (%) | ROUGE-1 |
|-----------------------|---------------|--------------|--------------|---------------|
| LSA | 54.76 | 42.59 | 47.08 | 0.4854 |
| MMR ($\lambda=1.0$) | 83.34 | 54.66 | 65.69 | 0.5379 |

summarization. In addition, between generic LSA and query-based LSA, better results are obtained in generic LSA. On the other hand, regarding the fact that MMR is inherently a query-based method and LSA is inherently a generic method, the best results are respectively obtained in query-based and generic approaches.

IV. CONCLUSION

In this study, we investigated unsupervised broadcast news summarization. MMR and LSA are the most well-known unsupervised methods in automatic speech summarization that are widely used in summarization. We suggested several systems to perform generic and query-based automatic broadcast news summarization that was based on Maximal Marginal Relevance (MMR) and Latent Semantic Analysis (LSA). The results demonstrated that MMR is preferred to LSA for query-based summarization, and LSA is preferred to MMR for generic summarization.

REFERENCES

- [1] E. Lloret and M. Palomar, "Text summarisation in progress: a literature review," *Artif. Intell. Rev.*, vol. 37, no. 1, pp. 1–41, 2012, doi: 10.1007/s10462-011-9216-z.
- [2] H. Saggion and T. Poibeau, "Automatic text summarization: Past, present and future," in *Multi-source, Multilingual Information Extraction and Summarization*, Springer, 2013, pp. 3–21.
- [3] M. Ramezani and M.-R. Feizi-Derakhshi, "Automated text summarization: An overview," *Appl. Artif. Intell.*, vol. 28, no. 2, pp. 178–215, 2014, doi: 10.1080/08839514.2014.862783.
- [4] Y. K. Meena, P. Dewaliya, and D. Gopalani, "Optimal Features Set for Extractive Automatic Text Summarization," in *Advanced Computing & Communication Technologies (ACCT), 2015 Fifth International Conference on*, 2015, pp. 35–40, doi: 10.1109/ACCT.2015.123.
- [5] M. Ramezani and M.-R. Feizi-Derakhshi, "Ontology-based automatic text summarization using FarsNet," *Adv. Comput. Sci. an Int. J.*, vol. 4, no. 2, pp. 88–96, 2015.
- [6] M. Ramezani and M.-R. Feizi-Derakhshi, "Achieving More Coherent Summaries in Automatic Text Summarization; an Ontology-based Approach," *J. Adv. Math. Comput. Sci.*, pp. 1–15, 2016, doi: 10.9734/BJMCS/2016/27549.
- [7] P. Sethi, S. Sonawane, S. Khanwalker, and R. B. Keskar, "Automatic text summarization of news articles," in *2017 International Conference on Big Data, IoT and Data Science (BIGD)*, 2017, pp. 23–29, doi: 10.1109/BIGD.2017.8336568.
- [8] M.-Y. Day and C.-Y. Chen, "Artificial intelligence for automatic text summarization," in *2018 IEEE International Conference on Information Reuse and Integration (IRI)*, 2018, pp. 478–484, doi: 10.1109/IRI.2018.00076.
- [9] J. N. Madhuri and R. G. Kumar, "Extractive text summarization using sentence ranking," in *2019 International Conference on Data Science and Communication (IconDSC)*, 2019, pp. 1–3, doi: 10.1109/IconDSC.2019.8817040.
- [10] W. S. El-Kassas, C. R. Salama, A. A. Rafea, and H. K. Mohamed, "EdgeSum: Graph-based framework for automatic text summarization," *Inf. Process. Manag.*, vol. 57, no. 6, p. 102264, 2020, doi: 10.1016/j.ipm.2020.102264.
- [11] C. Zong, R. Xia, and J. Zhang, "Automatic Text Summarization," in *Text Data Mining*, Springer, 2021, pp. 285–333.
- [12] G. Yasmin, S. Debnath, and A. K. Das, "Automatic Speech Summarization Without Linguistic Knowledge Based on Frame Selection Through Acoustic Features," in *Computational Intelligence in Pattern Recognition*, Springer, 2022, pp. 463–470.
- [13] S.-H. Liu, K.-Y. Chen, B. Chen, H.-M. Wang, and W.-L. Hsu, "Leveraging manifold learning for extractive broadcast news summarization," in *2017 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, 2017, pp. 5805–5809, doi: 10.1109/ICASSP.2017.7953269.
- [14] M. V. M. Cirne and H. Pedrini, "VISCOM: A robust video summarization approach using color co-occurrence matrices," *Multimed. Tools Appl.*, vol. 77, no. 1, pp. 857–875, 2018, doi: 10.1007/s11042-016-4300-7.
- [15] M. Elfeki and A. Borji, "Video summarization via actionness ranking," in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2019, pp. 754–763, doi: 10.1109/WACV.2019.00085.
- [16] S.-H. Liu, K.-Y. Chen, and B. Chen, "Enhanced language modeling with proximity and sentence relatedness information for extractive broadcast news summarization," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 19, no. 3, pp. 1–19, 2020, doi: 10.1145/3377407.
- [17] V. Tiwari and C. Bhatnagar, "A survey of recent work on video summarization: approaches and techniques," *Multimed. Tools Appl.*, pp. 1–35, 2021, doi: 10.1007/s11042-021-10977-y.
- [18] H. H. Kim, "Social speech summarization," *Proc. Am. Soc. Inf. Sci. Technol.*, vol. 48, no. 1, pp. 1–4, 2011.
- [19] A. Inoue, T. Mikami, and Y. Yamashita, "Improvement of speech summarization using prosodic information," in *Speech Prosody 2004, International Conference*, 2004, pp. 0–3.
- [20] S. Banerjee and A. I. Rudnicky, "An extractive-summarization baseline for the automatic detection of noteworthy utterances in multi-party human-human dialog," in *Spoken Language Technology Workshop, 2008. SLT 2008. IEEE*, 2008, pp. 177–180, doi: 10.1109/SLT.2008.4777869.
- [21] S. R. Maskey, "Automatic broadcast news speech summarization," 2008.
- [22] J. Zhang, R. R. H. Y. Chan, P. Fung, and L. Cao, "A comparative study on speech summarization of broadcast news and lecture speech," in *INTERSPEECH*, 2007, pp. 2781–2784, Accessed: Oct. 05, 2013. [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.152.5325&rep=rep1&type=pdf>.
- [23] H. Lee, Y. Chou, Y.-B. Wang, and L.-S. Lee, "Supervised Spoken Document Summarization jointly Considering Utterance Importance and Redundancy by Structured Support Vector Machine," 2012.
- [24] F. Liu and Y. Liu, "Exploring speaker characteristics for meeting summarization," in *INTERSPEECH*, 2010, pp. 2518–2521.
- [25] L. Wang and C. Cardie, "Domain-Independent Abstract Generation for Focused Meeting Summarization," in *ACL (1)*, 2013, pp. 1395–1405.
- [26] I. Zhang, H. Yuan, and X. Pan, "Rhetorical-state SVM for Lecture Speech Summarization," *Inf. Technol. J.*, vol. 13, no. 7, pp. 1447–1450, 2014, doi: 10.3923/itj.2014.1447.1450.
- [27] N. Garg, B. Favre, K. Reidhammer, and D. Hakkani Tür, "ClusterRank: a graph based method for meeting summarization," *Idiap*, 2009.
- [28] R. Mihalcea and P. Tarau, "TextRank: Bringing order into texts," 2004.
- [29] S. K. Jauhar, Y.-N. Chen, and F. Metz, "Prosody-Based Unsupervised Speech Summarization with Two-Layer Mutually Reinforced Random Walk," 2013.
- [30] Y.-N. Chen, Y. Huang, C.-F. Yeh, and L.-S. Lee, "Spoken Lecture Summarization by Random Walk over a Graph Constructed with Automatically Extracted Key Terms," in *INTERSPEECH*, 2011, pp.

- [31] H. Lee *et al.*, “Spoken knowledge organization by semantic structuring and a prototype course lecture system for personalized learning,” *IEEE/ACM Trans. Audio, Speech Lang. Process.*, vol. 22, no. 5, pp. 883–898, 2014, doi: 10.1109/TASLP.2014.2310993.
- [32] S.-H. Liu, K.-Y. Chen, B. Chen, H.-M. Wang, H.-C. Yen, and W.-L. Hsu, “Combining Relevance Language Modeling and Clarity Measure for Extractive Speech Summarization,” *Audio, Speech, Lang. Process. IEEE/ACM Trans.*, vol. 23, no. 6, pp. 957–969, 2015, doi: 10.1109/TASLP.2015.2414820.
- [33] D. Gillick, K. Riedhammer, B. Favre, D. Hakkani-t, and D. Hakkani-Tur, “A global optimization framework for meeting summarization,” in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, 2009, no. 2, pp. 4769–4772, doi: 10.1109/ICASSP.2009.4960697.
- [34] J. Carbonell and J. Goldstein, “The use of MMR, diversity-based reranking for reordering documents and producing summaries,” *Proc. 21st Annu. Int. ACM SIGIR Conf. Res. Dev. Inf. Retr. - SIGIR '98*, pp. 335–336, 1998, doi: 10.1145/290941.291025.
- [35] K. Zechner, “Automatic summarization of spoken dialogues in unrestricted domains,” Google, 2001.
- [36] K. Riedhammer, B. Favre, D. Hakkani-t, and D. Hakkani-Tur, “A keyphrase based approach to interactive meeting summarization,” in *Spoken Language Technology Workshop, 2008. SLT 2008. IEEE*, 2008, pp. 153–156, doi: 10.1109/SLT.2008.4777863.
- [37] S. C. Deerwester, G. W. Furnas, T. K. Landauer, R. A. Harshman, and S. T. Dumais, “Indexing by latent semantic analysis,” *JASIS*, vol. 41, no. 6, pp. 391–407, 1990, doi: 10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASII>3.0.CO;2-9.
- [38] Y. Gong and X. Liu, “Generic text summarization using relevance measure and latent semantic analysis,” in *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, 2001, pp. 19–25, doi: 10.1145/383952.383955.
- [39] Y. Yang and L. Xie, “Subword latent semantic analysis for texttiling-based automatic story segmentation of chinese broadcast news,” in *Chinese Spoken Language Processing, 2008. ISCSLP'08. 6th International Symposium on*, 2008, vol. 00, no. c, pp. 1–4, doi: 10.1109/CHINSL.2008.ECP.101.
- [40] Z. A. K. Aleqili and T. P. Tan, “Unsupervised Identification of Story Boundaries in Malay Spoken Broadcast News,” *J. Emerg. Technol. Web Intell.*, vol. 5, no. 1, 2013.
- [41] T. Hofmann, “Probabilistic latent semantic indexing,” in *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, 1999, pp. 50–57, doi: 10.1145/312624.312649.
- [42] R. M. Alguliev, R. M. Aliguliyev, M. S. Hajirahimova, and C. a. Mehdiyev, “MCMR: Maximum coverage and minimum redundant text summarization model,” *Expert Syst. Appl.*, vol. 38, no. 12, pp. 14514–14522, Nov. 2011, doi: 10.1016/j.eswa.2011.05.033.
- [43] A. Nenkova and K. R. McKeown, *Automatic summarization*, vol. 5, no. 2. Now Publishers Inc, 2011.
- [44] T. K. Landauer, S. T. Dumais, and S. T. Dumais, “A solution to Plato’s problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge,” *Psychol. Rev.*, vol. 104, no. 2, pp. 211–240, 1997, doi: 10.1037/0033-295X.104.2.211.
- [45] T. K. Landauer, D. S. McNamara, S. Dennis, and W. Kintsch, *Handbook of latent semantic analysis*. Psychology Press, 2013.
- [46] A. Nenkova and K. McKeown, “A survey of text summarization techniques,” in *Mining Text Data*, C. C. Aggarwal and C. Zhai, Eds. Boston, MA: Springer, 2012, pp. 43–76.
- [47] J. Steinberger and K. Jezek, “Using latent semantic analysis in text summarization and summary evaluation,” in *Proc. ISIM'04*, 2004, pp. 93–100.
- [48] G. Murray, S. Renals, and J. Carletta, “Extractive summarization of meeting recordings,” 2005, [Online]. Available: http://www.isc-speech.org/archive/interspeech_2005. J. Clerk Maxwell, *A Treatise on Electricity and Magnetism*, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–7