CmpE561 Research Project Report
Text Summarization with Latent Semantic Analysis

Mine Melodi Çalışkan 2015705009 *
Serkan Duman 2016700039 †

4 May 2017

Introduction

In this project automatic text summarization using Latent Semantic Analysis will be discussed which measures a content similarity between an original document and its summary.

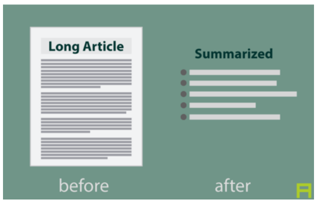
Automatic text summarization can be useful in a various fields in which huge amount of electronic documents makes it difficult to obtain the necessary information. It is an active research area and there are many applications currently used, e.g to summarize patience medical data for doctors is introduced by [17] , Reddit announced multimedia news summarizer "auto tl-dr" to summarize online texts and articles using an algorithm called "SMMRY" [19], and as stated in [18] summarization is useful for collating search engine hits to provide summary of the top N hits by a search.

Concept of the text summarization will be briefly introduced and Latent Semantic Analysis for summarization with it's derivation, including the topic of Singular Value Decomposition and it's use cases with the avaliable tools,datasets and it's state-of-the-art success rates following the evaluation methods will be covered.

* minemelodicaliskan@gmail.com
† serkan.duman@gmail.com

Text Summarization

Automatic Text summarization is a mathematical approach to make a computer program create a representative summary of the given document by finding the most significant sentences.



Input: A text document

Output: A subset of salient sentences from the document

Types of text summarizations

Extractive Document Summarizers

These methods extract salient sentences in the original text without modifying them to create a summary.

Abstractive Document Summarizers

These methods build an "internal" semantic representation and then creates summary using natural language generation tecniques. These summaries might contain words that are not explicitly present in the original text.

Statistical Summarizers

These methods use statistical features of the sentences ,e.g title, location, term frequency, assign weights to the keywords and then calculate scores of the sentences and select the highest scored sentence into the summaries.

1 Latent Semantic Analysis

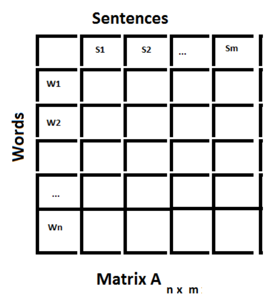
Latent Semantic Analysis (LSA) is an unsupervised summarization technique that uses Singular Value Decomposition (SVD) on the word-sentence matrix to extract semantically similar words and sentences. It's objective is to discover hidden semantic structures of words and sentences using context of the input document.

Input for this method is a words by sentences matrix A, where the columns represents sentences and the rows represents the words with the cells filled out representing the importance of words in sentences using different approaches (e.g number of occurence) which are discussed in the following sections.After calculation of SVD, sentence selection as a summary of the given document is performed on resulting singular value matrix.

Objective of LSA is to discover hidden semantic structures of words and sentences using context of the input document. We can imagine LSA performs SVD to "squeeze down" the input data matrix to lower rank by merging terms with similar co-occurence patterns to make "superterms" and to seek "latent" semantic associations from relationships between vectors using inner product in this lower dimensional space.

1.1 Filling the Input Matrix

In the representation of words × sentences matrix columns correspond to sentences and rows correnspond to words. The cell values are filled to represent the importance of the word in the sentences. Different approaches for calculation of these values are as follows:



1.1.1 Number of Occurrence

Frequency of the word in the sentence.

1.1.2 Binary Representation

If a word occurs in the sentence 1, otherwise 0.

1.1.3 Root Type

If the root type of the word is Noun, cell value is the frequency of the word, otherwise 0.

1.1.4 TF-IDF (Term Frequency-Inverse Document Frequency)

The importance of a word is high if it is frequent in the sentence, but less frequent in the document. TF-IDF is equal to TF×IDF

$$TF = tf(i,j) = \frac{\text{Frequency of word i in sentence j}}{\text{Sum of frequencies of all words in sentence j}}$$

$$IDF = idf(i,j) = log(\frac{\text{Number of sentences in input text}}{\text{Number of sentences containing word i}})$$

1.1.5 Modified TF-IDF

If cell values are less than or equal to the to the average TF-IDF values in the associated row set them to zero. This modification is done to eliminate noise effects.

1.1.6 Log-Entropy

Values of the cells are determined by log-entropy of the words i.e the amount of information in each sentence which is calculated as follows:

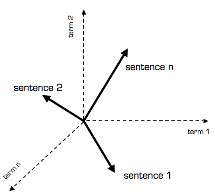
$$LogEntropy = (1 + \frac{\sum PlogP)}{logn}) * log(1 + f)$$

where

P :the probability of word i appeared in sentence j
f : the number of times word i appeared in sentence j
n : the number of sentences in the document

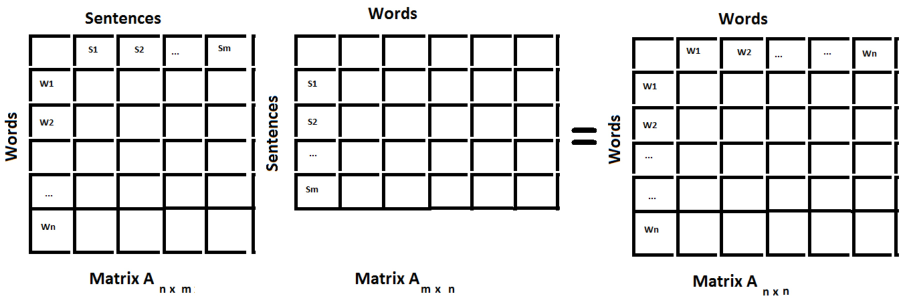
1.2 Sentences as points in the word space

In the words × sentences representation, sentences could be thought of as points in the word space. Dimensions of the word space correspond to various words and coordinates of the sentence are determined by e.g. the number of occurrences of the particular word in the sentence.

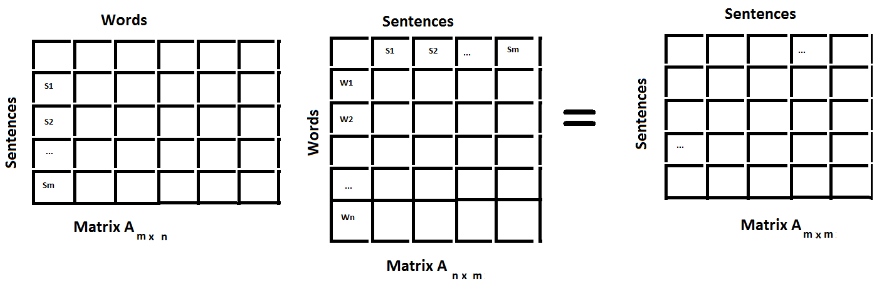


1.3 Word Similarity-Sentence Similarity Matrices

Let $\mathbb{W}$ be a $n \times n$ word similarity matrix $\mathbb{W} = AA^T$. In the binary-valued model, element $w_{ij} = \alpha_i \ \alpha_j$ represents number of sentences in which word i and word j co-occur , i.e appear together.

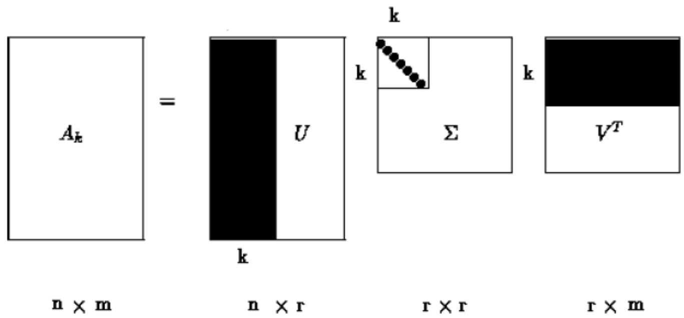


Let $\mathbb{S}$ be a $m \times m$ sentence similarity matrix $\mathbb{S} = A^T A$. In the binary-valued model, element $s_{ij} = \beta_i \ \beta_j$ represents a number of distinct words sentence i and sentence j have in common.



1.4 Singular Value Decomposition

Singular Value Decomposition is a data dimensionality reduction technique.It provides an exact representatition for the input data matrix as a product of three matrices and allows to eliminate less important parts of the data that are linearly independent to produce a much smaller matrix that approximate it with the desired number of dimensions. This approximation is called as "Low-rank approximation". As expected, accuracy is directly proportional to the number of the dimensions we choose.



Singular Value Decomposition and Dimensionality Reduction

Related theorem of SVD and the procedure of Low Rank approximation are as follows:

Theorem: Let r be the rank of the $n \times m$ matrix A. Then there is a singular value decomposition of A of the form:

$$A = U\Sigma V^T \tag{1}$$

where

- The eigenvalues λ_i , $i = 1, \ldots, r$ of AA^T are the same as the eigenvalues of $A^T A$

- Let $\sigma_i = \sqrt{\lambda_i}$ with $\lambda_i \geq \lambda_{i+1}$, where $i = 1, \ldots, r$. Then $n \times m$ matrix Σ is constructed by setting $\Sigma_{ii} = \sigma_i$, $i = 1, \ldots, r$ and zero otherwise. Here the values σ_i are the singular values of A.

An example of calculation of SVD can be found in Appendix B.

Procedure of Low-rank Approximation

- Given A, construct its SVD in the form as: $A = U\Sigma V^T$

- Derive from Σ the matrix Σ_k formed by replacing by zeros the r-k smallest singular values on the diagonal of Σ.

- Compute and output $A_k = U\Sigma_k V^T$ as the rank-k approximation to A.

Remark. This procedure yields the marix of rank k with lowest possible Frobenius error. Related theorem about this statement is discussed in Appendix A.

1.5 Singular Value Decomposition in LSA

LSA performs SVD on A to obtain the singular value matrix and select top k sentences as a summary of a given document. In this method, the given input matrix A decomposed into three matrices U, Σ and V^T where
U : Words $\times$ Concepts matrix
Σ : Scaling values, diagonal descending matrix
V^T : Concepts $\times$ Sentences matrix

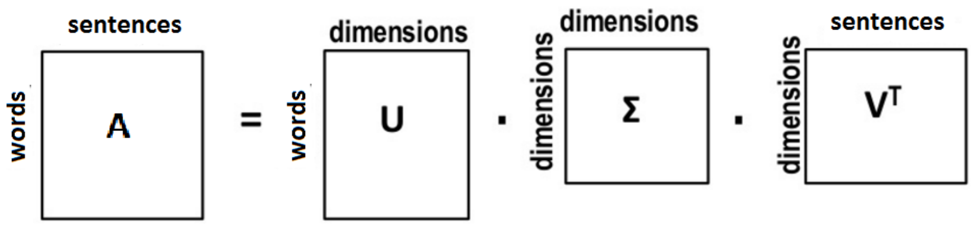
Procedure can be itemized as follows:

- Obtain SVD of A as:
  $A_{n\times m} = U_{n\times r}\Sigma_{r\times r}V_{r\times m}$

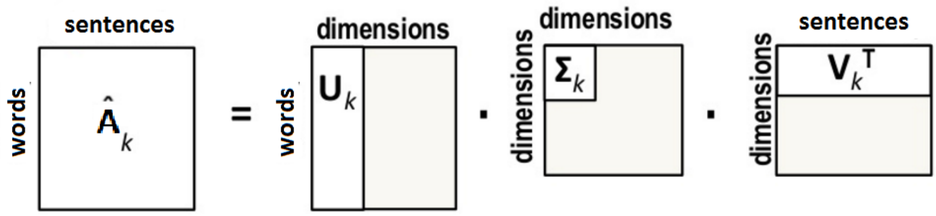
- Keep only k eigen values from Σ

- Approximate A with reduced dimensionality : $A_{n\times m} \sim U_{n\times k}\Sigma_{k\times k}V_{k\times m}$

- Convert words and sentences to points in k-dimensional space



SVD of the matrix X where U is the word eigenvector matrix, V is the sentence eigenvector matrix, $\sum$ is a diagonal matrix of singular values



Dimensionality Reduction

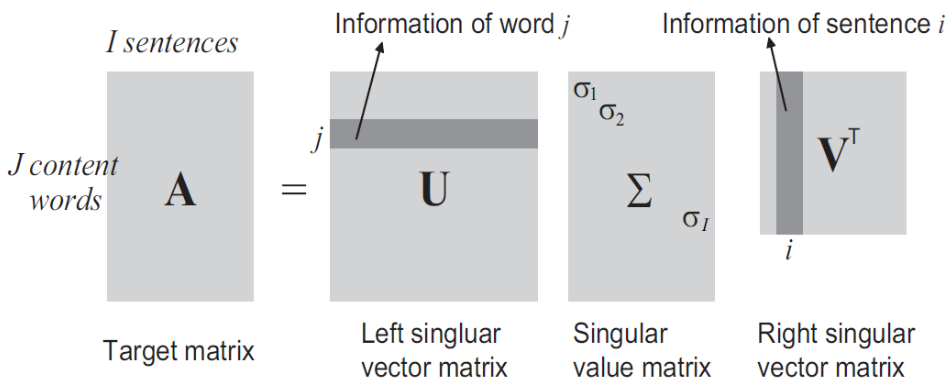
1.5.1 Meaning of columns and rows of U,V

The columns of $U_{n\times r}$ are the r eigenvectors of the $n \times n$ word similarity matrix $\mathbb{W}$,

$$\mathbb{W} = AA^T = U\Sigma V^T V\Sigma U^T = U\Sigma^2 U^T \tag{2}$$

The columns of $V_{m\times r}$ are the r eigenvectors of the $m \times m$ sentence similarity matrix $\mathbb{S}$,

$$\mathbb{S} = A^T A = V\Sigma U^T U\Sigma V^T = V\Sigma^2 V^T \tag{3}$$



Singular Value Decomposition for words × sentences matrix

1.6 Sentence Selection

1.6.1 Using V^T matrix, the matrix of concepts × sentences:

In V^T matrix row order shows the degree of the importance of the concepts, e.g the first row represents the most important concept extracted. The relation between the sentence and the concept is proportional to the cell values of V^T. The summarization process chooses the most informative sentence for each word. It means that the kth sentence we choose has the largest index value in kth right singular vector in matrix V^T.[9]
Steps of this approach is as follows:

- Choose one sentence from the most important concept.
- Then choose a second sentence from the second most important concept.
- Continue until desired number of sentences are collected.

Disadvantage: In this approach only one sentece from each concept is chosen. However, some important concepts may contain sentences that are highly related to the concept, but do not have the highest cell value.

1.6.2 Using both V and Σ matrices:

In this approach, length of each sentence vector, represented by the row of V matrix, is used for sentence selection. The length of the sentence i is calculated using the words whose indexes are less than or equal to the given dimension. Σ matrix is used as a multiplication parameter in order to give more emphasis on the most important words. The sentence with the highest length value is chosen to be a part of the resulting summary.[10]

Advantage: In this approach sentences that are related to all important concepts are chosen, while allowing collection of more than one sentence from an important concept.

1.6.3 Using V^T and Σ matrices:

In this approach, more than one sentence can be chosen from most important words, even if they do not have the highest cell value in the row of the related word and the reduced dimension does not have to be the same as the number of sentences in the resulting summary.[11]
Steps of this approach are as follows:

- Calculate the percentage of the related singular value over the sum of all singular values, for each concept using Σ matrix.
- Collect multiple sentences from each word using the provided result.

1.7 Simulation

We applied the algorithm to various texts [15, 16, 7] using number of occurence weighting scheme for the input matrix, and both V^T and Σ matrices for the summarization. The algorithm is applied after removing stop words to increase the performance and to avoid noises. Simulation results are provided in additional report paper.

2 System and Tools Currently Used

- **sumy 0.6.0 :** Module for automatic summarization of text documents and HTML pages. Available at $https://pypi.python.org/pypi/sumy$
- **SVDPlag v1.0 :** This tool is an example of automatic plagiarism detection system. It uses LSA framework to perform statistics computations which includes usage of SVD to infer the associations among

the common N-grams contained in the examined documents. Available at download section of $http://textmining.zcu.cz/$

3 Avaliable Data Sets and Corpora

In the work of Gong and Liu [9] as data corpus, two months of the CNN Worldview news programs are collected. Excluding commercial advertisements, a one day broadcast of this program lasts for about 22 mins, and on average, it consists of 15 individual news stories. The evaluation database consists of closed captions of 549 different news stories whose lengths are in the range of 3-105 sentences. Then all the stories with less than 10 sentences are eliminated for meaningful results of summarization, which results in 243 documents. As far as we know from our research, this dataset is not made publicly available. However, we saw a handful of researches by other academicians that used the same dataset for different tasks. Therefore we think that the dataset might be available upon request to the authors for academic purposes.

In [10], Steinberger and Jezek utilized Reuters Corpora (namely RCV1) and filtered out news documents with minimum length of 20 sentences. Reuters have released three large collections of news (RCV1, RCV2, and TRC2) by now, and NIST took over the distribution of readily available and any future Reuters Corpora. These datasets are available upon request to NIST for academic purposes.

In [11], Murray et al. used human summaries of the ICSI Meeting corpus which is collected from natural meetings that occurred at the International Computer Science Institute (ICSI) in Berkeley, California and contains simultaneously recorded audio and transcript of 75 meetings of 4 main types and 53 unique speakers. It is readily available on the URL [20]

In [12], Ozsoy et al. used two different sets of articles in Turkish, chosen from various areas such as medicine, sociology, psychology, where each set consists of 50 articles. Second data set contains longer articles. The abtract of these articles are utilized to compare with the automatic summaries. They didn't share any information about the availability of their corpus.

4 The Success Criteria used for Evaluation

Evaluation of automatic text summaries is by itself a very challenging and active research area. To tackle this problem, several approaches are developed:

4.1 General Approaches for Summary Evaluation

4.1.1 Evaluation by Sentence Co-selection

Co-selection measures include precision and recall of co-selected sentences. This method requires having a "right summary" to which we could compute precision (P) and recall (R) along with F-measure. The most common way to obtain such summaries is utilizing human summarizers for the documents of interest and declaring the average of these summaries as "ideal (right) summaries". This is problematic in various ways. Firstly, it is time-consuming, and secondly, it is subjective since human summarizers may take highly different approaches in the summarization process.

4.1.2 Content-based Methods

To clear out the weaknesses of co-selection measures, content-based similarity measures (e.g. cosine similarity) between a full text document with its summary are utilized. For instance to content-based similarity measure, cosine similarity can be computed by the following formula:

$$\cos\left(X,Y\right) = \frac{\Sigma x_i * y_i}{\sqrt{\Sigma (x_i)^2 * \Sigma (y_i)^2}} \tag{4}$$

where X and Y are the representations of full text and its summary on the vector space model, respectively.

4.1.3 Relevance Correlation

Relevance correlation is a measure for accessing the relative decrease in retrieval performance when indexing summaries instead of full documents.

4.1.4 Task-based Evaluations

Task-based evaluations measure human performance using the summaries for a certain task after the summaries are created. We can for example measure a suitability of using summaries instead of full texts for text categorization. This evaluation method requires a classified corpus of texts.

In the precursor work of Gong and Liu [9] where they proposed a novel method for text summarization based on LSA, they simply used manual summarization results to evaluate the performance of the LSA summarizer. They employed three human evaluators where each evaluator is requested to select exactly 5 sentences from each document, which he/she deemed the most important for summarizing the story. They also asked for consensus to have a combined result determined by majority votes to summarize each of the documents since the disagreements between the human summarizers were much more than expected, where in average 9 candidate and only 2.5 overlapping sentences are selected per document.

4.2 LSA-Based Evaluation Methods

To resolve the problems in the sentence co-selection method used in [9], Steinberger and Jezek proposed two novel LSA-based evaluation methods that are content-based evaluation approaches. They measure cosine similarity between an original document and its summary using the U matrices produced in the SVD process. The two proposed measures are:

4.2.1 Similarity of the main topic

$$\cos\varphi = \sum_{i=1}^{n} u_i^e * u_i^f \tag{5}$$

, where u^f is the first left singular vector (i.e. the main topic) of the full text SVD, and u^e is the first left singular vector of the summary SVD, n is the number of unique terms in the full text. These first left singular vectors correspond to the main topic (most salience word pattern) of full text and its summary.

4.2.2 Similarity of the term significance

This method compares a summary with the full document from an angle of n major topics, and has the following advantage to the previous one. Suppose, a document contains two topics with relatively same significance. When the second one outweighs the first one in a summary, the main topic of the summary will not be consistent with the main topic of the full text. This method takes more singular vectors into account and removes this weakness.

$$s_k = \sqrt{\sum_{i=1}^{n} u_{k,i}^2 * \sigma_i^2} \tag{6}$$

, where s_k is the length of the k'st term vector in the modified latent vector space, n is the number of dimensions of the new space. (See [9] for details)

4.3 ROUGE Evaluation Approach

This method is based on n-gram co-occurance, longest common subsequence and weighted longest common subsequence between the ideal summary and the extracted summary. Details of this evaluation algorithm can be found in [13].

5 State-of-the-art Success Rates

As explained in section 4, success rate is dependent on the evaluation methods being used. For the methods based on LSA, weighing scheme of term-frequencies by sentences is also a crucial factor for the success rate.

In [9], Gong and Liu tied the dimensionality reduction to summary lenght. Also in their approach, a sentence that has very high correspondance to a concept might lose to a less important sentence for the sake of reduced redundancy due to its rank in its own concept, or the concept it belongs to is not within best k concepts even it's strong, since the summary consisted of k sentences that have 1-to-1 relation with k most important concepts in the full text. They evaluated the performance of their summarization method with reference to manual summaries, which can only be generated by following a time consuming and highly subjective process. Their method reached F-score of F=0.57.

In [10], to remedy the weak points of Gong and Liu approach pointed out previously, Steinberger and Jezek proposed a method of sentence extraction with variable dimensionality reduction (meaning variable summary lenght) along with two LSA-based evaluation methods (see section 4.2). The aim of their algorithm is to choose sentences which are related to all important concepts rather than being strongly related to just one concept and at the same time to be able to choose more than one sentence from an important topic.

In [11], Murray et al. also suggested an approach where one or more sentences are collected from the topmost concepts in V^T matrix. The number of sentences to be selected depends on the values in the Σ matrix. The

Table 1: ROUGE-L Scores for the data set 1

	G&L	S&J	MRC	Cross	Topic
frequency	0,236	0,250	0,244	0,302	0,244
binary	0,272	0,275	0,274	0,313	0,274
tf-idf	0,200	0,218	0,213	0,304	0,213
log-entropy	0,230	0,250	0,235	0,302	0,235
root type	0,283	0,282	0,289	0,320	0,289
mod. tf-idf	0,195	0,221	0,223	0,290	0,223

Table 2: ROUGE-L Scores for the data set 2

	G&L	S&J	MRC	Cross	Topic
frequency	0,256	0,251	0,259	0,264	0,259
binary	0,191	0,220	0,189	0,274	0,189
tf-idf	0,230	0,235	0,227	0,266	0,227
log-entropy	0,267	0,245	0,268	0,267	0,268
root type	0,194	0,222	0,197	0,263	0,197
mod. tf-idf	0,234	0,239	0,232	0,268	0,232

percentage of the related singular value over the sum of all singular values determines the number of sentences to be collected from each topic.

In [12], Ozsoy et al. used ROUGE evaluation approach (Lin and Hovy, 2003) to measure the performance of two sentence selection strategy they proposed, namely Cross and Topic, and all the other methods mentioned in papers within this section, on their corpus that contains Turkish articles (see section 3). The results are as in Table 1 and Table 2, where G&L is Gong & Liu, S&J is Steinberger & Jezek, MRC is Murray, Renals & Carletta.

Observing the results in Table 1, Cross method outperforms the other methods for all matrix creation techniques. It also indicates that all algorithms give their best results when input matrix is created using the root type of words. Results within Table 2 also indicates that Cross has one of the best score, if not the best, for each weighing scheme amongst all algorithms. For this dataset, which simply contains longer Turkish articles than the first one, when the log-entropy method is used, all algorithms achieved their best F-score based on ROUGE.

6 Approaches for different languages

Weighted term-frequency by sentence matrix A is sparse by its nature. The sparseness of data is even more for documents in agglutinative languages such as Turkish, where it becomes less possible to produce any meaningful relations between terms in a sentence. To resolve this problem, root type of words or only noun terms are used to for text classification tasks.

In case of text summarization, given the evaluation results of text summarization methods used on Turkish articles in [12], best F-score is achieved for all compared methods when the root type method and log-entropy method is used for matrix creation, which reflects the same property.

As of now, there is not enough research to provide meaningful results on language specific features of weighing scheme used to construct matrix A, methods of sentence selection and performance evaluation between languages with such different structures such as English and Turkish.

7 Conclusion

Generic text summarization and its evaluation are very challenging research areas, since neither query nor topic are provided in the process. Therefore the summarization method, summaries and and their performance judgements tend to lack consensus. As both a summarization and a performance evaluation method, LSA based approaches suffer from this fundamental properties of generic summarization as other methods such as Maximal Marginal Relevance (i.e. MMR), Feature-based approaches.

Latent Semantic Analysis is a algebraic-statistical approach, and by its nature is an uninformed clustering algorithm where it clusters latent topics of a document, which simply reveals k most strong latent concepts (i.e. topics in our case) that are inferred only by the linear dependency of input dimensions (which are words) in the dataset. Appliyng various strategies on sentence selection, and various evaluation methods, there is still room to improve the performance of such generic text summarization approaches mentioned in this research. Ideas used in other methods such as graph based approaches, or future-based approaches may be used together with LSA to improve the performance of summarization.

A Low Rank Approximation

Let A be a data matrix, and let $A = U\Sigma V^T$ be the SVD of A, where U and V are orthogonal and Σ is diagonal.Then, for a given k, the optimal rank k approximation of A is given by

$$best_k(A) = U_k \Sigma_k {V^T}_k \tag{7}$$

where U_k and V_k are the matrices formed by the first k columns of the matrices U and V , respectively, and Σ_k is the k-th principal submatrix of Σ.

Remark. This rank k approximation is the best possible approximation with respect to the Frobenius norm, among all matrices with rank k. [14] The error in this approximation is equal to σ_{k+1} .Thus the larger k is, the smaller this error and for $k = r$, the error is zero since $\Sigma_r = \Sigma$ provided $r < M, N$, then $\sigma_{r+1} = 0$ and thus $A_r = A$

B Computing the SVD of a Matrix

To calculate SVD of a matrix A, first we need to find eigenvalues and eigenvectors of AA^T and A^TA. Resulting eigenvectors of AA^T and A^TA will be columns of U and V, respectively and square roots of eigenvalues from AA^T (or A^TA) will be the singular values of Σ.

Example: Let matrix A be given as:

$$A = \begin{bmatrix} 2 & 4 \\ 1 & 3 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Calculation of AA^T

$$AA^T = \begin{bmatrix} 2 & 4 \\ 1 & 3 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & 4 & 0 & 0 \\ 1 & 3 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 20 & 14 & 0 & 0 \\ 14 & 10 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$(AA^T - \lambda I)x = \begin{bmatrix} 20-\lambda & 14 & 0 & 0 \\ 14 & 10-\lambda & 0 & 0 \\ 0 & 0 & -\lambda & 0 \\ 0 & 0 & 0 & -\lambda \end{bmatrix}$$

Setting $det(AA^T - \lambda I) = 0$ we have:

$\Rightarrow \lambda_1 = 0, \lambda_2 = 0, \lambda_3 \sim 29.883, \lambda_4 \sim 0.117$

Substituting the eigenvalue λ_3 we obtain:

$$\begin{aligned} -9.833x_1 + 14x_2 &= 0 \\ 14x_1 - 19.883x_2 &= 0 \\ x_3 &= 0 \\ x_4 &= 0 \end{aligned} \tag{8}$$

$\Rightarrow x_1 = -0.58, x_2 = 0.82$ and $x_3 = x_4 = 0$ is the second column of the matrix U.

And substituting λ_4

$$\begin{aligned} 19.833x_1 + 14x_2 &= 0 \\ 14x_1 + 9.833x_2 &= 0 \\ x_3 &= 0 \\ x_4 &= 0 \end{aligned} \tag{9}$$

$\Rightarrow x_1 = 0.82, x_2 = -0.58 \text{and} x_3 = x_4 = 0$ is the first column of the matrix U.

Combining these results we obtain:

$$U = \begin{bmatrix} 0.82 & -0.58 & 0 & 0 \\ 0.58 & 0.82 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Similarly computing

$$A^T A = \begin{bmatrix} 2 & 4 & 0 & 0 \\ 1 & 3 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & 4 \\ 1 & 3 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

and doing similar analysis to find V we obtain:

$$V = \begin{bmatrix} 0.40 & -0.91 \\ 0.91 & 0.40 \end{bmatrix}$$

Finally we take square root of the eigenvalues (from AA^T or A^TA) to find Σ

$$\Sigma = \begin{bmatrix} 5.47 & 0 \\ 0 & 0.37 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Remark. $\sigma_1 > \sigma_2$

Resulting singular value decomposition of the matrix A is:

$$A = \begin{bmatrix} 2 & 4 \\ 1 & 3 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0.82 & -0.58 & 0 & 0 \\ 0.58 & 0.82 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 5.47 & 0 \\ 0 & 0.37 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0.40 & 0.91 \\ -0.91 & 0.40 \end{bmatrix} = U\Sigma V^T$$

References

[1] Makoto Hirohata, Yousuke Shinnaka, Koji Iwano and Sadaoki Furui *Sentence Extraction-based Presentation Summarization Techniques And Evaluation Metrics* Department of Computer Science, Tokyo Institute of Technology

[2] Bob Glushko *Dimensionality Reduction and Latent Semantic Analysis*

[3] Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze *Introduction to Information Retrieval* Cambridge University Press. 2008.

[4] Finny G Kuruvilla, Peter J Park, Stuart L Schreiber *Vector atalgebra in the analysis of genome-wide expression data* Genome Biology

[5] Jieping Ye *Generalized Low Rank Approximations of Matrices* Department of Computer Science, University of Minnesota, Minneapolis, MN 55455, USA

[6] Desmond Jordan,Gregory Whalen, Blaine Bell, Kathleen McKeown, Steven Feiner

An Evaluation of Automatically Generated Briefings of Patient Status

[7] Fannie Flagg *Fried Green Tomatoes, Story About A Lake*

[8] Document Similarity in Information Retrieval Mausam Based on slides of W. Arms, Thomas Hofmann, Ata Kaban, Melanie Martin

[9] Gong, Y. and Liu, X. 2001. *Generic Text Summarization using Relevance Measure and Latent Semantic Analysis*

[10] Steinberger, J. and Jezek, K. 2004. *Using Latent Semantic Analysis in Text Summarization and Summary Evaluation*

[11] Murray, G., Renals, S. and Carletta, J. 2005. *Extractive Summarization of Meeting Recordings*

[12] Ozsoy, Makbule G., Cicekli, I. and Alpaslan, Ferda N. 2010. *Text Summarization of Turkish Texts using Latent Semantic Analysis*

[13] Lin, C.Y., and Hovy, E. 2003.
Automatic Evaluation of Summaries Using N-gram Co-occurence Statistics

[14] Eckart, Carl, and Gale Young 1936. *The approximation of a matrix by another of lower rank* Psychometrika 1: 211-218

[15] Pablo Neruda *Sonnet Lxxxi*

[16] Anne Sexton *The Truth the Dead Know*

[17] N. Elhadad and K.R. McKeown. *Towards Generating Patient Specific Summaries of Medical Articles* NAACL Automatic Summarization

[18] Dragomir R. Radev Weiguo Fan *Automatic summarization of search engine hit lists*

[19] Reddit
www.reddit.com/user/autotldr
http://smmry.com

[20] ICSI Corpus
http://www1.icsi.berkeley.edu/Speech/mr/

[21] Saeedeh Gholamrezazadeh, Mohsen Amini Salehi, Bahareh Gholamzadeh. *A Comprehensive Survey on Text Summarization Systems* 2009 Computer Science and its Applications, 2nd International Conference.