Text Summarization and Singular Value Decomposition

Conference Paper · January 1970

DOI: 10.1007/978-3-540-30198-1_25

2 authors:

Josef Steinberger
University of West Bohemia

Karel Jezek
University of West Bohemia

Text Summarization and Singular Value Decomposition

Karel Ježek[1], Josef Steinberger[1]

[1] University of West Bohemia in Pilsen,
Department of Computer Science and Engineering,
30614, Univerzitni 22, Plzeň, Czech Republic
{jstein, jezek_ka}@kiv.zcu.cz

Abstract. In this paper we present the usage of singular value decomposition (SVD) in text summarization. Firstly we mention the taxonomy of generic text summarization methods. Then we describe the principles of the SVD and its possibilities to identify semantically important parts of a text. We propose a modification of the SVD-based summarization, which improves the quality of generated extracts. In the second part we propose two new evaluation methods based on SVD, which measure content similarity between an original document and its summary. In evaluation part, our summarization approach is compared with 5 other available summarizers. For evaluation of a summary quality we used, apart from a classical content-based evaluator, both newly developed SVD-based evaluators. Finally, we study **the** influence of the summary length on its quality from the angle of the three evaluation methods mentioned.

1 Introduction

The actual huge amount of electronic information has to be reduced to enable the users to handle this information more effectively. Short summaries can be presented to users, for example, in place of full-length documents found by search engine in response to a user's query. In section 2 we mention prior approaches to text summarization and **section** 3 covers our recent research focus. In section 4 we describe **the** method based on SVD which has been recently published. We have further modified and improved this method. One of the most controversial fields in the summary research is its evaluation process. Next part of the article deals with possibilities of summary evaluation. We propose there two new evaluation methods based on SVD, which measure a content similarity between an original document and its summary. At the end of the paper we present evaluation results and further research directions.

2 Approaches in Automatic Text Summarization

We will now present a brief overview of prior work in the text summarization. We can begin with classical approaches that include the use of surface level indicators of informative relevance and corpus statistics that can be applied to unrestricted text. Luhn (1958) developed the first sentence extraction algorithm which uses term frequencies to measure sentence relevance [7]. Kupiec et al. (1995) implemented a trainable Bayesian classifier that computes the probability that a sentence in a source document should be included in a summary [5]. The next group consists of methods which take the text cohesion into account. An example is the lexical chains method which searches for chains of context words in the text [6]. Ono et al. (1994) and Marcu (1997) made use of Rhetorical Structure Theory, which is a descriptive theory about text organization, as the bases for text summarization. The approach consists in the construction of a rhetorical tree for a given text [8]. Knowledge intensive approaches are based on the extensive encoding of world knowledge about specific situations. These methods base the selection of information not on the surface level properties of the text, but on expected information about a well known situation. The next approach is mapping natural language into predefined, structured representations, that**,** when instantiated**,** represent the key information from the original source (e. g. Concept-based abstracting – Jones and Paice, 1992, [9]). While sentence extraction is a currently wide-spread and useful technique, more research in summarization **is now** moving towards summarization by generation. Jing and McKeown (2000) proposed a *cut-and-paste* strategy as a computational process of automatic abstracting and a sentence reduction strategy in order to produce concise sentences [10]. A quite new approach in text summarization uses the singular value decomposition.

3 Our Previous Summarization Research

Our recent research has been focused namely on the use of inductive machine learning methods for automatic document summarization. We analyzed various approaches to document summarization, using some existing algorithms and combining these with a novel use of itemsets. The resulted summarizer was evaluated by comparing classification of original documents and that of **a** summary generated automatically [3]. **Now/Then ?** we decided to investigate possibilities of using singular value decomposition in both creating a summary and its evaluation.

4 SVD-based Summarization

Yihong Gong and Xin Liu have published the idea of using SVD in text summarization in 2002 [1]. The process starts with creation of a term by sentences matrix $\mathbf{A} = [A_1, A_2, \ldots, A_n]$ with each column vector A_i, representing the weighted term-frequency vector of sentence i in the document under consideration. If there are a total of m terms and n sentences in the document, then we will have an $m \times n$ matrix $\mathbf{A}$ for the document. Since every word does not normally appear in each sentence, the matrix $\mathbf{A}$ is sparse.

Given an $m \times n$ matrix **A**, where without loss of generality $m \geq n$, the SVD of **A** is defined as:

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

where $\mathbf{U} = [u_{ij}]$ is an $m \times n$ column-orthonormal matrix whose columns are called left singular vectors; $\mathbf{\Sigma} = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n)$ is an $n \times n$ diagonal matrix, whose diagonal elements are non-negative singular values sorted in descending order, and $\mathbf{V} = [v_{ij}]$ is an $n \times n$ orthonormal matrix, whose columns are called right singular vectors (see figure 1). If rank(**A**) = r, then (see [4]) **Σ** satisfies:

$$\sigma_1 \geq \sigma_2 \ldots \geq \sigma_r > \sigma_{r+1} = \ldots = \sigma_n = 0\,. \tag{2}$$

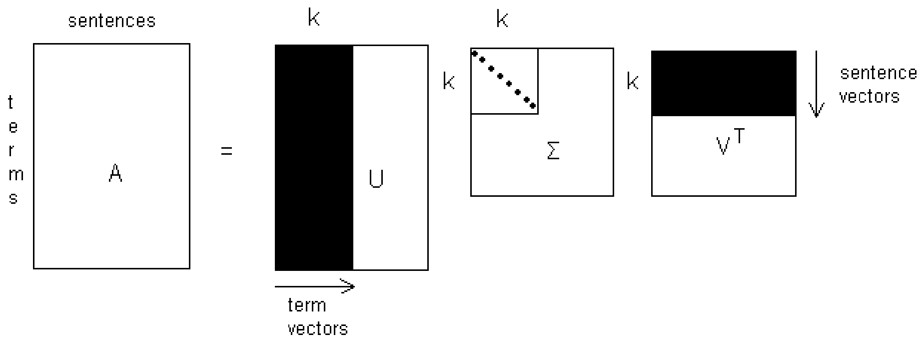

Fig. 1. Singular Value Decomposition

The interpretation of applying the SVD to the terms by sentences matrix **A** can be made from two different viewpoints. From transformation point of view, the SVD derives a mapping between the m-dimensional space spawned by the weighted term-frequency vectors and the r-dimensional singular vector space. From semantic point of view, the SVD derives the latent semantic structure from the document represented by matrix **A**. This operation reflects a breakdown of the original document into r linearly-independent base vectors or concepts. Each term and sentence from the document is jointly indexed by these base vectors/concepts. A unique SVD feature is that it is capable of capturing and modelling interrelationships among terms so that it can semantically cluster terms and sentences. **Furthermore**, as demonstrated in [4], if a word combination pattern is salient and recurring in document, this pattern will be captured and represented by one of the singular vectors. The magnitude of the corresponding singular value indicates the importance degree of this pattern within the document. Any sentences containing this word combination pattern will be projected along this singular vector, and the sentence that best represents this pattern will have the largest index value with this vector. As each particular word combination pattern describes a certain topic/concept in the document, the facts described above naturally lead to the hypothesis that each singular vector represents a salient topic/concept of the document, and the magnitude of its corresponding singular value represents the degree of importance of the salient topic/concept.

Based on the above discussion, authors [1] proposed a summarization method which uses the matrix $\mathbf{V}^T$. This matrix describes an importance degree of each topic in each sentence. The summarization process chooses the most informative sentence for each topic. It means that the k'th sentence we **choose/chose?** has the largest index value in k'th right singular vector in matrix $\mathbf{V}^T$.

5 Modified SVD-based Summarization

The above described summarization method has two significant disadvantages. At first it is necessary to use the same number of dimensions as is the number of sentences we want to choose for a summary. However, the higher_ the number of dimensions of reduced space, the less significant topic we take into a summary. This disadvantage turns into an advantage only in the case when we know how many different topics_ the original document **has** and we choose the same number of sentences into a summary. The second disadvantage is that a sentence with large index values, but not the largest (it doesn't win in any dimension), will not be chosen although its content _ for the summary **is** very suitable.

In order to clear out the discussed disadvantages**,** we propose following modifications in the SVD-based summarization method. Again we need to compute SVD of a term by sentences matrix. We get the three matrices as **shown in figure** 1. For each sentence vector in matrix $\mathbf{V}$ (its components are multiplied by corresponding singular values) we compute its length. The reason of the multiplication is to favo**u**r the index values in the matrix $\mathbf{V}$ that correspond to the highest singular values (the most significant topics). Formally:

$$s_k = \sqrt{\sum_{i=1}^{n} v_{k,i}^2 \cdot \sigma_i^2}\,, \qquad (3)$$

where s_k is the length of the vector of k'th sentence in the modified latent vector space. It is its significance score for summarization too. n is a number of dimensions of the new space. This value is independent **of** the number of summary sentences (it is a parameter of the method). In our experiments we chose the dimensions whose singular values didn't fall under the half of the highest singular value (but it is possible to set a different strategy). Finally, we put into **a/the** summary the sentences with the highest values in vector s.

6 Summary Evaluation Approaches

Evaluation of automatic summarization in a standard and inexpensive way is a difficult task. It is **an** equally important area as the own summarization process and that's why many evaluation approaches were developed [2].

Co-selection measures include precision and recall of co-selected sentences. These methods require having at **one's** disposal **a/the** "right extract" (to which we could compute precision and recall). We can obtain this extract in several ways. The most common way is to obtain some human (manual) extracts and to declare the average of these extracts as **the** "ideal (right) extract". However, obtaining_ human extracts is usually problematic. Another problem is that two manual summaries of the same input do not share in general many identical sentences.

We can clear out the above discussed weakness of co-selection measures by content-based similarity measures. These methods compute the similarity between two documents at a more fine-grained level than just sentences. The basic method evaluates the similarity between the full-text document and its summary with the **cosine** similarity measure, computed by the following formula:

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

where X and Y are representations based on the vector space model.

Relevance correlation is a measure for accessing the relative decrease in retrieval performance when indexing summaries instead of full documents [2].

Task-based evaluations measure human performance using the summaries for a certain task (*after* the summaries are created). We can for example measure **the** suitability of using summaries instead of full-texts for text categorization [3]. This evaluation requires a classified corpus of texts.

7 Using SVD in Summary Evaluation

We classify this new evaluation method to a content-based category because, like the classical cosine content-based approach (see 6.), it evaluates a summary quality via content similarity between a full-text and its summary. Our method uses SVD of the terms by sentences matrix (see 4.), exactly the matrix **U**. This matrix represents the degree of importance of terms in salient topics/concepts. In evaluation we measure the similarity between the matrix **U** derived from the SVD performed on the original document and the matrix **U** derived from the SVD performed on the summary. For appraising this similarity we have proposed two measures.

7.1 First Left Singular Vector Similarity

This method compares first left singular vectors of the full-text SVD (i. e. SVD performed on the original document) and the summary SVD (i. e. SVD performed on the summary). These vectors correspond to the most salient word pattern in the full-text and its summary (we can call it the main topic).

Then we measure the angle between the first left singular vectors. They are normalized, so we can use the following formula:

$$\cos\varphi = \sum_{i=1}^{n} ue_i \cdot uf_i \,, \tag{5}$$

where *uf* is the first left singular vector of the full-text SVD, *ue* is the first left singular vector of the summary SVD (values, which correspond to particular terms, are sorted up the full-text terms and instead of missing terms are zeroes), *n* is a number of unique terms in the full-text.

7.2 U.Σ-based Similarity

This evaluation method compares a summary with the original document from an angle of *n* most salient topics. We propose the following process:

- Perform the SVD on a document matrix (see 4.).
- For each term vector in matrix **U** (its components are multiplied by corresponding singular values) compute its length. The reason of the multiplication is to favour the index values in the matrix **U** that correspond to the highest singular values (the most significant topics). Formally:

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

  where s_k is the length of the *k*'st term vector in the modified latent vector space, *n* is a number of dimensions of the new space. In our experiments we chose the dimensions whose singular values didn't fall under the half of the highest singular value (by analogy to the summary method described above).
- From the lengths of the term vectors (s_k) make a resulting term vector, whose index values hold an information about the term significance in the modified latent space (see figure 2).
- Normalize the resulting vector.

This process is performed on the original document and on its summary (for the same number of dimensions according to the summary) (see figure 2). In the result, we get one vector corresponding to the term vector lengths of the full-text and one of its summary. As a similarity measure we use again the angle between resulting vectors (see 7.1).

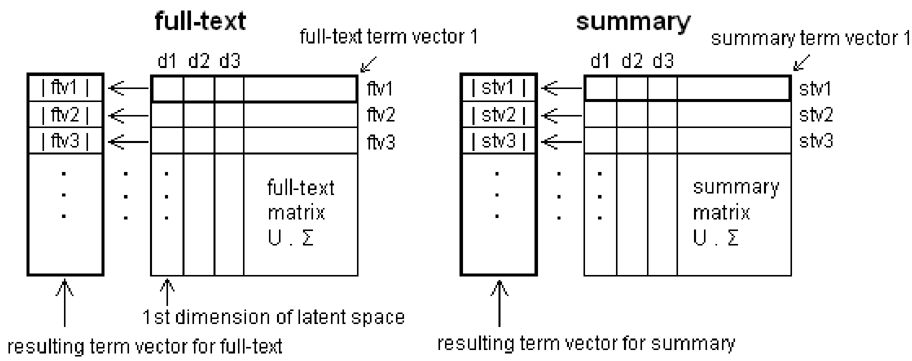

Fig. 2. Creation of a resulting term vectors of a full-text and a summary

This evaluation method has the following advantage above the previous one. Suppose, an original document contains two topics with **virtually the same?** significance (corresponding singular values are almost the same). When the second significant topic outweighs the first one in a summary, the main topic of the summary will not be consistent with the main topic of the original. Taking more singular vectors (than just one) into account removes this weakness.

8 Experiments

8.1 Testing Collection

We tested our document summarizer using the Reuters Corpus Volume 1 (RCV1) collection (the first "official" collection Reuters corpus released to the community of researches, containing over 800 thousand documents). We prepared a collection by selecting RCV1 documents with **the** length **of** at least 20 sentences. The selected documents had to be suitable for the summarization task. Table 1 contains details about our collection.

Table 1. Testing collection – details

Number of documents	127
Minimum number of sentences in document	20
Maximum number of sentences in document	68
Average number of sentences per document	28
Average number of words per document	724
Average number of significant words per document	287
Average number of distinct significant words per document	187

8.2 Results and Discussion

We evaluated the following summarizers:

- Gong + Liu SVD summarizer (SVD–G+L)
- SVD summarizer based on our approach (SVD–OUR)
- RANDOM – evaluation based on the average of 10 random extracts
- LEAD – first *n* sentences
- 1-ITEMSET – based on itemsets method [3]
- TF.IDF –based on frequency method [3]

These summarizers were evaluated by the following three evaluation methods:

- Cosine similarity – classical content-based method
- SVD similarity – First left singular vector similarity
- SVD similarity – U.Σ-based similarity

The summarization ratio was set to 20 %. Results are presented in the following table. Values are averages of cosines of angles between a full-text and its summary.

Table 2. Summary quality evaluation

Evaluator	Summarizer					
	SVD-L+G	SVD-OUR	RANDOM	LEAD	1-ITEMSET	TF.IDF
Cosine similarity	0,761	**0,765**	0,663	0,753	0,759	0,753
First left sing. vector simil.	0,751	**0,787**	0,488	0,73	0,764	0,758
U.Σ-based similarity	0,824	**0,851**	0,542	0,771	0,817	0,803

The classical cosine evaluator shows only small differences between summarizers (the best summarizer – 0,77 and the worst (random) - 0,65). It **is** caused by a shallow level of this evaluation method which takes into account only term counts in compared documents. The evaluation based on SVD is a more fine-grained approach. It is possible to say that it evaluates a summary via term co-occurrences in sentences.

Figures 3-5 show the dependencies of a summary quality on the summarization ratio and the evaluation methods for our SVD-based and random summarizer.

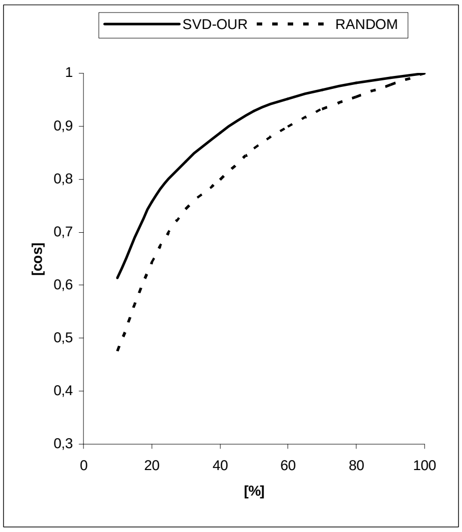

Fig. 3. Cosine Similarity Evaluation

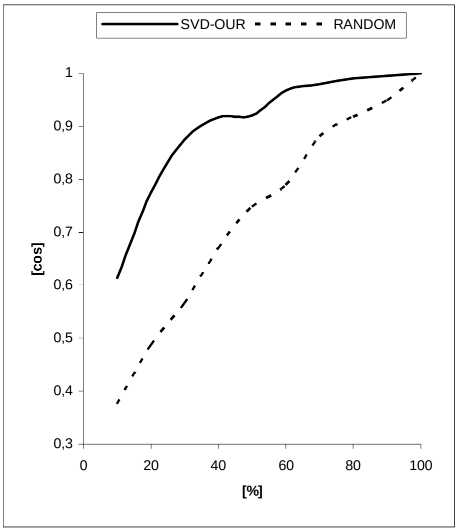

Fig. 4. First singular vector similarity evaluation

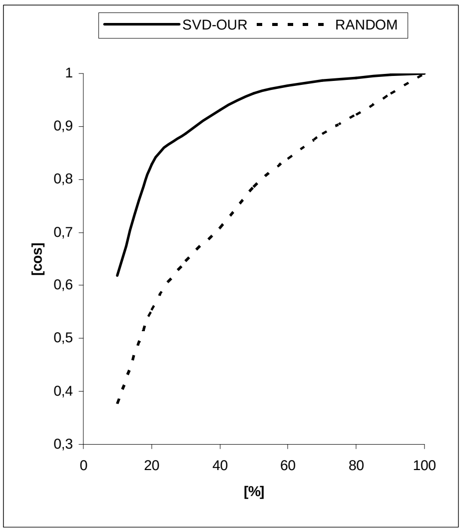

Fig. 5. U.Σ-based Similarity Evaluation

In the evaluation by the first left singular vector we noticed the disadvantage discussed in 6. (proved in 10% of documents). The U.Σ-based evaluation removes this weakness. There is also a big difference between random and other summarizers. **Next** we observed from the evaluation, that the SVD summarizer has been shown as the expressively best with the evaluator (3). This property was expected.

9 Conclusion

This paper introduced a new approach to automatic text summarization and summary evaluation. The practical tests proved that our summarizing method outperforms the other examined methods. Our other experiments showed that SVD is very sensitive on a **stoplis**t and a **lemmatization** process. Therefore we are working on improved versions of lemmatizers for English and Czech languages. In future research we plan to try other weighing schemes and a normalization of a sentence vector on the SVD input. Of course, other evaluations are needed, especially on longer texts than the Reuters documents are. Our final goal is to integrate our summarizer in to a natural language processing system capable of searching and presenting web documents in a concise and coherent form.

This work has been partly supported by grants No. MSM 235200005 and ME494.

References

1. Gong, Y., Liu, X.: Generic Text Summarization Using Relevance Measure and Latent Semantic Analysis. Proceedings of the 24th ACM SIGIR conference on Research and development in information retrieval, New Orleans, Louisiana, United States (2001) 19-25
2. Radev, R., Teufel, S., Saggion, H., Lam, W., Blitzer, J., Qi, H., Celebi, A., Liu, D., Drabek, E.: Evaluation Challenges in Large-scale Document Summarization. Proceeding of the 41st meeting of the Association for Computational Linguistics, Sapporo, Japan (2003) 375-382
3. Hynek, J., Ježek, K.: Practical Approach to Automatic Text Summarization. Proceedings of the ELPUB ’03 conference, Guimaraes, Portugal (2003) 378-388
4. Berry, M.W., Dumais, S.T., O’Brien, G.W.: Using Linear Algebra for Intelligent Information Retrieval. SIAM Review (1995)
5. Kupiec, J., Pedersen, J., Chen, F.: A trainable Document Summarizer. Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval, Seattle, Washington, United States (1995) 68-73
6. Barzilay, R., Elhadad, M.: Using Lexical Chains for Text Summarization. Proceedings of the Intelligent Scalable Text Summarization Workshop (ISTS'97), ACL Madrid, Spain (1997)
7. H. P. Luhn: Automatic Creation of Literature Abstracts. IBM Journal and Research Development 2(2) (1958) 159-165
8. Marcu D.: From Discourse Structures to Text Summaries. Preceedings of the ACL’97/EACL’97 Workshop on Intelligent Scalable Text Summarization, Madrid, Spain (1997) 82-88
9. Jones, P.A., Paice, C.D: A ‘select and generate’ Approach to Automatic Abstracting. Proceeding of the 14th British Computer Society Information Retrieval Colloquium, Springer Verlag (1992) 151-154
10. Jing, H., McKeown, K.: Cut and Paste Text Summarization. Proceedings of the 1st meeting of the North Americat Chapter of the Association for Computational Linguistics, Seatle, Washington, USA (2000) 178-185