Journal of Theoretical and Applied Information Technology
30th April 2019. Vol.97. No 8
© 2005 – ongoing JATIT & LLS
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

SEMANTIC VARIANCE MEASUREMENTS FOR EVALUATING SINGLE-DOCUMENT SUMMARIZATION

EIMAN TAMAH AL-SHAMMARI

Department of Information Sciences, College of Computing Sciences and Engineering, Kuwait University, Kuwait

E-mail: dr.eiman@kuniv.edu.kw

ABSTRACT

This paper proposes a new measurement for evaluating single document summarizers using a novel formulation of Latent Semantic Analysis techniques . Additionally, we introduce a measure to the *conceptual coherence* of a passage against a document or set of documents, based on singular value decompositions of term-document matrices. The proposed work aims to reflect the coverage concepts in a passage by quantifying the significance of semantic variance of documents and words, per concept in the entire concept space. Our method developed emphasize on completeness, rather than the soundness of a summary. The performance based on the initial tests on a variety of corpus types shows promising results that comply with human evaluation.
Keywords: *LSA, Summarization, Evaluation, Semantic Variance*, *SVD, Term-document Matrices.*

1 INTRODUCTION

1.1 Summarization and Evaluation

Text Summarization is a beneficial tool for interpreting textual information and helps in highlighting the most important information in no time [1]. Moreover, it's difficult for human beings to manually condense large documents. Hence the goal of text summarization is to shorten the information content by preserving the overall meaning of the document. Text Summarization is a compressed version of a text or set of texts carrying the most salient of concepts, the most vital pieces of information communicated in the text, stated in no more than half the length of the original text(s), and usually far less than that in length [2] . Hence the problem of summarization involves identification of what is 'important', what is novel, what represents shared and/or similar information, how to compose it in alternate form, and so on.

Text summarization being the most challenging tasks has extensive application areas [3, 4]. As the internet grows the amount of information is abundant making it difficult to select relevant information. The possible uses of summarization includes multimedia news summarization, text to speech for blind people, collating search engine hints, summarizing meetings, producing intelligence reports, multiple document and multimedia summarization etc.

The methods of automatic text summarization include extraction based summarization and abstraction based summarization. In the extractive process the summarizing agent (be it human or any other intelligent agent) makes decisions on what to include or exclude in a compressed digest of the text corpus, often dealing with complex issues of linguistic synonymy and style. Whereas abstractive process uses linguistic methods to interpret the text by extracting the text in form of abstract which is later processed to obtain the original meaning of the text document.

The abstract statement of the problem makes the evaluation of summarization performance equally difficult (Kallimani,2012). Every summarization algorithm presents an implied notion of what a good summary for a particular instance should be, that in turn becomes an intrinsic standard for evaluation. That is, for some established summarization procedure S, a summary for an instance of textual input can simply be evaluated in terms of the ideal produced by S, by some quantitative metric.

Evaluating quality of text summary is a very important task. Text summarization must be evaluated with appropriate methods and the quality of summary can be compared with the

© 2005 – ongoing JATIT & LLS
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

system summary or human generated summary or summary generated by other system.

Summarization evaluation methods can be classified in to two categories (Spark, Galliers 1995), intrinsic and extrinsic evaluation. Intrinsic summarization evaluation measures the system by itself often done by some standard reference summarization systems or by using man-made informatics. Intrinsic evaluation focusses on the coherence and informativeness of the summaries.

Extrinsic evaluation methods also exist (Lin et al. 2006, Lin and Hovy 2002, Lin et al. 2004), tend to depend on the manual (human) construction of ideal summaries, often as summary segments or units, which are then used as reference bases for a heuristic comparison with a candidate summary.

But again, how many ideal summaries can be driven from various humans? Is the ideal summary unique?

Having an evaluation method that does not depend on human agents at any stage in the evaluation, thus being completely automated is very much desirable for reasons of efficiency (cost, effort) and scale. The value of such a method would be reliant on its general compliance with human evaluation of summary/gist production.

In this paper we aim to provide such a method, utilizing linear algebraic methods from LSA in identifying prevalent concepts. The methods developed emphasize ***completeness***, rather than the soundness of a summary; we present tools to evaluate the extractive performance of an agent. This is as opposed to the measurement of more linguistically-oriented abstractive abilities of a summarizing agent.

2 RELATED WORK

Automatic text summarization is an active research area from early 1950's. An important research in (Luhn 1958) analysed the information from word frequency and measuring their relative significance for words, sentences and then extracting accurate abstraction of the document. The result of information overload on the Web lot of approaches and evaluation methods have been explored in recent years. A survey of summarization and summary evaluation techniques can be found in (Das, Martins 2007, Goldstein 2005).

The proposed work in (Sarkar, Kamal. 2013) uses key concepts identified from a document for creating the summary of document. The text summarization with novel use of key concepts uses position based sentence filtering to eliminate less important sentences and efficiently takes care of redundancy which is a critical issue in text summarization. Latent semantic analysis LSA is a technique for extracting the relational meanings of terms, sentences or documents on the basis of their contextual use and latent semantic indexing LSI is an indexing and retrieval method that uses mathematical technique called singular value decomposition matrix containing word count to identify relations between terms and concepts in a document. As noted in (Dumais 1995) uses a reduced dimension vector space to represent words and documents and was similarly devised to improve recall rather than precision. Text summarization using LSA (Ozsoy et al. 2011) presents an algebraic statistical method to extract hidden semantic structures of words and sentences and provides an extractive Turkish text summarization based on LSA. (Steinberger et al. 2004) uses latent semantic text analysis in text summarization which measures the content similarities between an original document and the summary. Another approach used for text summarization is the use of Recall-Oriented Understudy for Gisting Evaluation (ROUGE) (Lin 2004) is a software package for automatic analysis of summaries. The evaluation approach can automatically decide on the quality of a summary by making a comparative study with other ideal summaries designed by humans. Text segmentation can be defined as a processes of breaking documents in to coherent mutli-paragraph subparts such as words, sentences and topics and represents the topics that are included in a text and how the meaning of topic change with in a text. Use of segmentation aids in obtaining better summaries in text summarization. (Dias et al. 2007).

presents a context based topic segmentation based on new informative similarity measures based on word co-occurrence. The main contribution the new weighting schemes and the definition of new similarity measures proposes a mathematical model to deal with co-occurrence factor and avoids an extra step of boundary detection. The proposed work in (Angheluta et al. 2002) presents use of topic segmentation in automatic summarization which a valuable first step for summarizing informative text. It yields good summaries in the form of TOCs and

Journal of Theoretical and Applied Information Technology
30th April 2019. Vol.97. No 8
© 2005 – ongoing JATIT & LLS
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

acceptable summaries in single and multiple documents. (Kamble et al. 2017)

2.1 LSA

Latent Semantic Analysis (LSA) is an algebraic statistical method that extracts meaning of words and similarity of sentences using the information about the usage of the words in the context. The more common words between sentences mean that those sentences are more semantically related. LSA method can represent the meaning of words and the meaning of sentences simultaneously. It averages the meaning of words that a sentence contains to find out the meaning of that sentence. It represents the meaning of words by averaging the meaning of sentences that contain this word.

LSA method is built on Singular Value Decomposition Matrix (SVD), a mathematical decomposition technique closely similar to factor analysis which models relationship among words and sentences. It has the capability of noise reduction, which leads to an improvement in accuracy.

The advantages of using LSA include that it selects the word importance count from the information provided by the corpus, it sums up semantic similarity between words which broadens the equivalent relation between words. Although LSA has advantages it has few limitations that it uses only the information in the input text, and it does not use the information of world knowledge, it does not use the information of word order, syntactic relations, or morphologies such information is used for finding out the meaning of words and texts and the performance of the algorithm decreases with large and inhomogeneous data. The decrease in performance is observed as SVD's complex algorithm is used for finding out the

Latent Semantic Analysis (also called Latent Semantic Indexing in Information Retrieval literature) introduced in (Deerwester et al. 1990), intends to present words and documents in abstract semantic terms as vector spaces, by decomposing a term-document matrix built from a text corpus. It keeps information about which words are used in a sentence, while preserving information of common words among sentences.

LSA applies the Singular Value Decomposition (SVD) for text summarization. SVD is used for finding out semantically similar words and sentences. SVD is a method that models relationships among words and sentences and derives the latent semantic structure from the document. A feature of SVD includes that it is capable of picking and designing interrelationships among words that can semantically gather words and sentences.

We provide a brief overview here of the general method. Initially a typical term-document frequency matrix A is constructed: For a corpus of m words and n documents, an $m \times n$ matrix A is produced. A decomposition

$$⟦SVD = f: A \rightarrow U, \sum, V⟧ \hat{} t \quad (1)$$

produces three matrices $⟦U, V⟧ \hat{} t$ and $\sum$ of dimensions $m \times z$, $z \times z$ and $z \times n$, for $z = \min(m; n)$ such that:

$$⟦A = U \sum V⟧ \hat{} t \quad (2)$$

This is called a Singular Value Decomposition. $\sum \in \mathbb{R} \hat{} (z \times z)$ is a diagonal matrix.

Since the number of documents n is usually less than the number of terms m, it is normally the case that $z = n$, therefore. $\sum \in \mathbb{R} \hat{} (n \times n)$. The terms in $\sum$ are non-negative, non-increasing order, generally thought of as representing the scalar strength of 'concepts' in the space of the documents.

The SVD is the first phase in LSA procedures, as it casts the words against the 'concepts' rather than the documents, in the vector space of matrix product

$$U \sum \in \mathbb{R} \hat{} (m \times n) \quad (3)$$

It similarly allows documents to be represented against derived concepts in the product

$$\sum V \hat{} t \in \mathbb{R} \hat{} (n \times n) \quad (4)$$

This is the simple basis for further LSA work, which usually involves some rank reduction in the number of concepts by truncating singular value matrix $\sum$. Distance similarity metrics between words and documents, or words and words, or documents and each other, can then be computed.

A new matrix A may also be created from the product of truncated U, $V \hat{} t$ and $\sum$ to represent word-document relations without 'insignificant' concepts, and this can have non-zero entries for words that do not appear in a

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

document, through second-order relations deduced by the SVD and expounded by the $\sum$ reduction. The dimensions of reduction is as follows : In matrix $\sum$, k largest values are selected by keeping corresponding values of U $and\, V$^t The resulting matrix A_k is given by

$$A_k = U_k \Sigma_k V^T_k \quad (5)$$

where k, k<z is the dimensionality of the concept space. The k selected must be large enough to allow applicable characteristics of the data and small enough to filter irrelevant details.

A more detailed explanation of and set of examples can be found in the original literature (Deerwester et al. 1990) and subsequent efforts along this path. Applying distance/similarity computations, the LSA computations have been used in a variety of domains (Geiss 2011) .

2.2 Textual coherence with LSA

LSA provides a fully automatic method for comparing entire textual information to each other to create their semantic relatedness and is used as an approach to measure the coherence of text (Khorsi et al. 2018). The text data is compared to each other using a derived measures of their sematic similarities. The measure can be based on mathematical analysis of relations among words and passages in a large corpus. Measure of semantic relatedness is similar to the measure of coherence because it captures the precise semantically related information that reflect semantic similarity among synonyms, antonyms, hyponyms etc., in the text document. (Foltz et al. 1998) discusses the measurement of textual coherence with latent semantic analysis. The general approaches discussed utilises LSA to perform coherence predictions to compare units of data in text to determine the degree of semantic relation between them. LSA coherence analysis require two text units to compare their sematic relatedness. A large text such as sentences, passages can be represented as vectors in space and each text unit corresponds to weighted average of vectors of the terms it contains in the document. The semantic relatedness of two text unit can then be compared by deriving the cosine between the two vectors and hence the coherence between the two text units can be determined. For example, two sentences in a text unit that uses exactly same terms with same frequencies can have cosine of 1, sentences containing some terms with related meaning but are not same may have moderate cosines, and sentences which are not semantically related may have cosine of zero or even below.

2.3 Passage-to-Document Coherence

Coherence indicates to the "joining together" property of the text i.e., it is the degree to which chunks of information (words, terms) throughout the document are related and can be linked. Passages are shorter than the documents and hence are more coherent (focused) units, they can aid in creating summaries that are very "informative" than those summaries generated from whole document.

Given a single document D consisting of a set of passages$(d_0, d_1, d_2, \ldots, d_n\,)$, the one-to-many (passage to document) relation is as follows: for a singular value decomposition $⟦U, \sum, V⟧$ ^t of the term-document matrix A, and subsequently produced 2D concept spaces $\alpha_{0..k}$ and $\beta_{0..k}$ where

$$\boldsymbol{\alpha} = ⟦\textstyle\sum \boldsymbol{V}⟧ \text{ \textasciicircum}\boldsymbol{t} \quad (6)$$

and

$$\boldsymbol{\beta} = U\textstyle\sum \quad (7)$$

we compute the semantic variance C for a document passage or text-block d_i as

$$\boldsymbol{C} = \textstyle\sum_{m=0}^{k} |\boldsymbol{\alpha_i}|\ \boldsymbol{\sigma_m} \quad (8)$$

where σ_m is the standard deviation in vector m for all d_j ≠ d_I in the corpus. This is our basic formula. We may write it more explicitly:

$$\boldsymbol{C} = \textstyle\sum_{m=0}^{k} |\alpha_{i,m}|\ \sigma_{\alpha(m)} \quad (9)$$

where $\sigma\_\alpha(m)$ is the square root of the variance around zero in vector space α for a concept m and all passages $d\_j \in\ d\_0, d\_1, d\_2, \ldots, d\_n$, $j{\neq}i$. It is a measure of the variance in semantic significance of the concept m to all documents other than the one being tested.
Since words and documents are mapped onto the same concept space, we may alternatively use the slightly different variance obtained from the word-concept vector, β_m, so that the equation becomes

$$\boldsymbol{C} = \textstyle\sum_{m=0}^{k} |\alpha_{i,m}|\ \sigma_{\beta(m)} \quad (10)$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

which is a somewhat more conservative variance measure according to the experi- ments with our implementation below (see section 3.2).

2.4 Inter-Passage Coherence

The formula above uses abosulte values of the concept-document vectors α to nul- lify the effects of direction. We posit that a large value of $\alpha\_(i, m)$ implies a significance of concept m to document passage d_i, regardless of the directionality of d in the vector space. Equivalently, a value close to zero implies the passage is not related to the concept, in the matrices decomposed this way.
The variance σ_α of all other documents (or the variance σ_β of all words in terms of that concept) is a positive variable factor that will be large for concepts expressed in many documents and small for concepts that are insignificant to most documents (and similarly to most words).
A separate measure for the shared semantic variance (i.e., the cross-concept coherence) of two passages x and y can be written as

$$\boldsymbol{CC} = \sum_{m=0}^{k} \frac{|\alpha_{x,m}\alpha_{y,m}|}{\sigma_{\alpha(m)}} \qquad (11)$$

where once again σ_m is the standard deviation around the zero point; $\sigma\_\beta(m)$ for words and $\sigma\_\alpha(m)$ for passages. Either version of σ may be used with slightly contrasting results.

2.5 Paper Overview

We introduce new measures in section 2.1.1, both for one-to-one and one-to-many relations. In 2.1.3 we discuss a basic adaptation of this as a summary evaluation technique, and perform a series of experiments on single-document document summarization, in 3.2 and 3.3.

3 METHODOLOGY

LSA methods for text content are usually based on some notion of distance measurements between entities (documents, parts of documents...etc.,) in the conceptual vector spaces produced by the singular value factoring process.

This work is built on top of basic linear algebraic techniques of LSA, which has found application in several domains. The notion of semantic variance utilizes the SVD matrix decompositions of LSA in a different direction, emphasizing cross-concept significance of text blocks other than the one concerned. In cross concept approach after the input matrix creation and the calculation of SVD, Vt matrix is used for sentence selection purpose. This measure provides a completely automated means of evaluating the one-to-many relationship of a textual passage to the remainder of the associated document corpus. We have shown the general compatibility of this score with a separate (also novel) measure for one-to-one conceptual coherence of two documents, and that both measures comply with human semantic evaluation.

Semantic Analysis is explored outside of natural language applications and applied to data from several fields like security and finance, where the idea of ‘concepts’ defined through first and second-order occurrence of symbols (in our textual case: words) has shown promise as a heuristic input to classification algorithms. The conceptual coherence measure here functions as a non-normalised similarity score and can be applied to several fields in a similar fashion. Transciptomics in biology research is of interest, having a vast amount of data with obscure, but important, inter-block relationships.

Euclidean distance and cosine similarity scores are predominant measures for this task. Distance similarity measures emphasize the idea that similar documents will be aligned in direction in the concept space. We propose that this may not always hold similar pieces of text may not always, in the vector spaces produced by singular value decomposition, have a small angle between them; they may not be generally ‘close’ to each other at all. Instead they may share a measurable significance across concepts that are independent of direction. We present a formulation of this using Inter-Passage Coherence.

In applying to the problem of document summarization, or more precisely the evaluation of summaries, we have produced term-document matrices from text corpora segmented into smaller mini-documents, or passages. All term document matrices used in our calculations in this literature are, therefore, word–passage representations.

4 EXPERIMENTATION

We employ the methods above in single document scenarios. For the single document procedure, the text is divided into its paragraphs (or similar length sections), and the term-

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

document matrix is constructed from these. There may be anywhere between one to several hundred passages in any one document, especially of the type concerned for summarization tasks.

A text pre-processing stage precedes the inclusion of terms in the construction of term-passage matrices. After the construction of the original (decomposable) matrix and subsequent decomposition, the variance and shared-variance measures on the summary passage are straightforward to calculate.

4.1 Pre-processing the Text

Pre-processing is precise representation of original text in a document. The goal of pre-processing is to eliminate overall outcome of sentences in a document that are related to concepts somehow but are not the core sentence for that concept. Text pre-processing and normalization is common for many NLP tasks including summarization. The bag-of-words approach throughout this paper requires a certain level of cleaning of the text, which we perform prior to normalization by lemmatization. Cleaning is a process of removing non-alphanumeric elements in the text and replacing them with whitespace, removing punctuation, and normalizing letter case. The pre-processing also include sentence boundary identification to identify the presence of dot at the end of sentence; stop word elimination in which common words with no meanings (semantics) and which do not combine relevant information to the summaries are eliminated. We specify a minimal set of standard stop-words to be removed, after which the remaining set of words from the text corpus may either be further reduced by lemmatization or stemming. The purpose of stemming is to obtain basis of each word which indicates its meaning.

In this paper, we perform no heuristic stemming of words. Instead we rely on a WordNet-based Lemmatizer (provided by Bird et al. 2009, Miller, 1995) that provides more efficient combination of traditional lexicographic data and modern computing it is an online lexical database created for use under program control. WordNet lemmatizer uses part-of-speech tagging to extract term lemmas of an input instance, from a large dictionary. This ensures the terms/lemmas ultimately processed by the algorithms are 'correct', i.e., exist in human vernacular. Terms not reducible to WordNet lemmas were used as is. All terms are delimited by whitespace.

4.1.1 Note on Lemmatization Effects

Lemmatization is a process of defining lemma (grouping together) different inflected form of words that can be analysed in to single term. Lemmatization operates on full text document and hence can distinguish between words that have different semantics.

Producing lemmas from words is still a heuristic process, even when based on a concrete dictionary-based method such as WordNet Lemmatizer. This is because while all the lemmas produced may be linguistically valid, they may not necessarily keep the intended meaning for a particular context. For instance, 'variability' and 'variable' may have very different intended meanings in an academic context, and so a summary including one word may not necessarily be reflecting information containing the other.

4.2 Single-Document Summary Evaluation Experimental setting

We created datasets from documents of various types. For the single-document scenario, each document presents a single dataset/corpus:

(A) Academic Paper Tom Mitchell's survey paper "Machine Learning" (Mitchell 2006). The paper is a medium length (approx. 5 page) survey on Machine Learning theory, current practices and results, and future direction as an interdisciplinary field of research. It contained about 840 unique terms after lemmatization, in 44 text segments.

(B) Academic Paper Our second academic paper is an in-depth technical/experimental paper by Joachims (Joachims 1997)

(C) Over 20 pages in length, it contained about 650 unique terms and was divided into 112 passages.

Summaries were created for each corpus, and experiments on their evaluation using our methods produced encouraging results.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

We used the basic semantic variance C-score (equation 7) as an indicator for how well a given passage relates to the corpus – the single document.

Additionally we applied Rouge-L (Longest Common Subsequence), to demonstrate the relevance of our summary evaluation analysis to the Rouge standardized evaluation tool.

The table below shows the results. Where we first would like to comment on our proposed evaluation technique, in accordance with intuition, a good summary will relate significantly to the major concepts prevalent in the document, and so produce a high normalized C-score. As for Rouge Results, as stated by (Lin and Hovya 2004)

Table 1: Normalized Passage-Document Coherence Scores and Rouge-L evaluation on the Single Document Texts

	Our Technique	ROUGE-L
Corpus A Nonsense Vs. document	0.077	0
Corpus A Human Vs. Document	0.953	0.556
Corpus B Nonsense Vs. Document	0.063	0
Corpus B Human Vs. Document	0.967	0.1

ROUGE package does not perform well with long documents, thus the values for Corpus B were not surprising. The fact that ROUGE requires these parameters indicates that you must not place too much reliance on its results - by adjusting the parameters you can get a range of results. This (in my opinion) compromises its usefulness. However, it is clear from each file that the human summary is a lot better than the rubbish summary.

Figures 1 and 2 provide a 2-dimensional illustration of the contributing semantic variance in concepts of low and high significance. We picked the SVD-derived concepts of highest and lowest significance to illustrate the effect on the semantic variance.

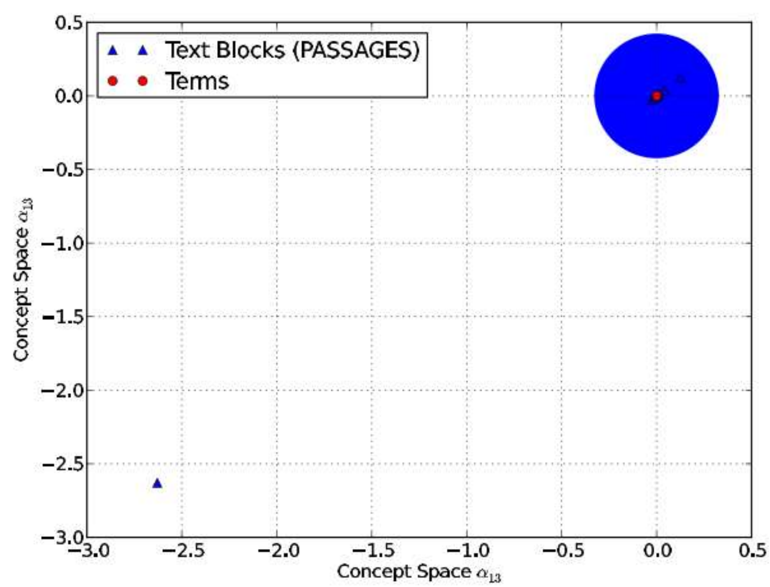

Figure 1: A plot of α and β spaces for non-significant concepts, from Single- Document Corpus A. A solitary passage shows relevance to this concept space; and the variance level here is low.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

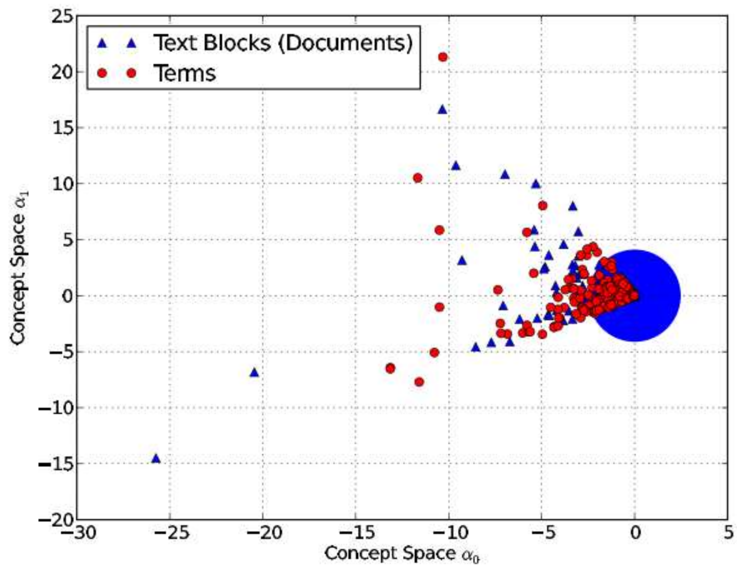

Figure 2: A plot of α and β spaces for significant concepts, from Single-Document Corpus A. Many passages show relevance (large absolute values) for the concepts, in alternate directions, illustrating a high variance level.

A more interesting case is the set of randomly extracted group of sentences from each corpus. This is also a summary, but most likely not a 'good' one. It relates to some passages and some concepts, but depending on how the sentences are picked, the summary evaluation is affected.

How 'good' such a summary is will depend on how likely it is for the sentence- extraction method to pick sentences that happen to contain significant concepts. This is related to (and possibly a function of, depending on concept distribution) the length of the document. It is easier to arbitrarily happen on significant sentences and words when the document is short (one page) than when the document contains hundreds of passages. This is illustrated in our preliminary table of results, particularly in the inaugural speech corpus.

To further elaborate on the utility of this score in its compliance with human evaluation of a single document summary, we conducted trial runs in which summaries of increasing randomness were tested. Figure 3 is a plot of these results.

Starting from the human expert summary we iteratively replace pieces of text with randomly picked phrases from the document, then again with randomly picked words that may not be in the text. The progression of score values shows the decreasing relevance of the summary passage to the document.

Figure 3 also shows the variability in the score at each level of randomized summary construction. For instance, a summary that is half-randomized (approx. level 5 in our plot) is more sensitive to variation than the manually produced ideal. Regenerating the randomized summary (i.e., re-seeding the random summary generator) may produce scores significantly lower or higher than the mean of summaries at that level.

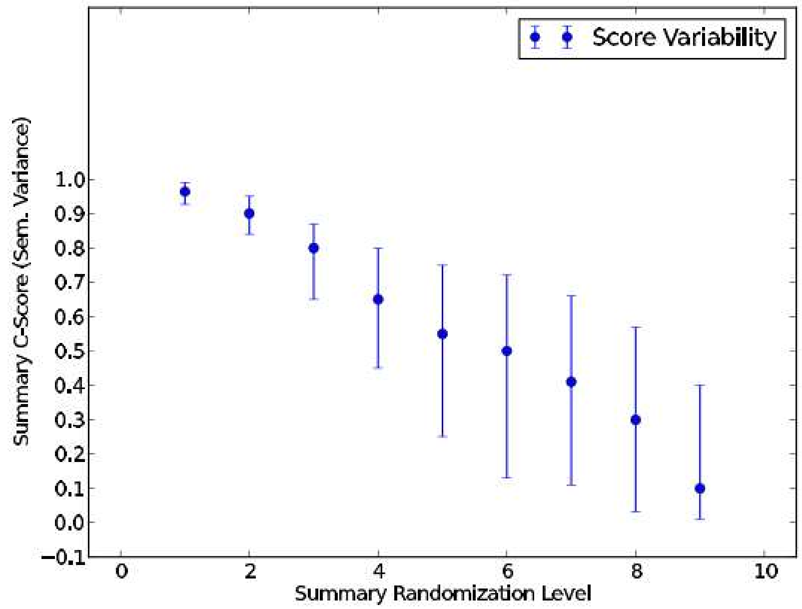

Figure 3: Summary performance, in decreasing levels of soundness. Illustrated also is the sensitivity of the scores to changes in the summary content at each level relevance to this concept space; and the variance level here is low.

Ideal gists vary in score as well (to a lesser degree) as illustrated, since there is no absolute ideal. Two different human agents can produce very different wordings for a gist or summary of a document. This is a general fact owing to subjectivity in human summarizing agents (see Kolluru and Gotoh) for an exploration of this in a dedicated work). In our method, different wordings of an 'ideal' summary sharing much of the same information will result in slightly contrasting C-scores.

5 CONCLUSIONS

The evaluation of a summary quality is an ambitious task. Automatically evaluating summaries is hard to accomplish as machines cannot comprehend the meaning of the summary , thus evaluating its quality. Needless to say having an "ideal summary" to compare to is an additional barrier as there is no unique ideal summary.

We did develop a method that emphasize on completeness, rather than the soundness of a summary.

To evaluate our technique, we used the basic semantic variance C-score as an indicator for how well a given passage relates to the corpus – the single document.

Additionally we compared our work to the state of the art Rouge-L, to demonstrate the relevance of our summary evaluation analysis to the Rouge standardized evaluation tool.

6 RESEARCH ASSUMPTOPNS AND FUTURE WORK

The initial set of tests described in this paper have demonstrated the applicability of the measures introduced, relying on the notion of semantic variance, to quantify passage-passage and passage-document relations in the space of semantic concepts, particularly in the framework of text summarization. We have specifically concerned ourselves with the evaluation of summarization performance, and these measures have so far proven useful.

An important issue that has been largely side-stepped throughout this work, is paragraphing – or document segmentation. It is imperative that the effects of para- graphing on the measures described be investigated for any applicative domain with the multitude of methods in this approach, because of the influence of para-graph selection on the division of concepts in any type of document or contiguous sequence of symbols.

For this work in the textual domain we have assumed normal (explicitly provided) paragraphing, usually groups of 2 to 5 sentences,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

as an acceptable division. A more finely grained segmentation of the single document may reflect on summary evaluation capabilities. We have already noted effects of document size and text-block size. Do single sentences hold enough of a 'concept' to be effective as 'blocks'? Indeed, whatever the choice of term and document in any field, similar exercises with paragraphing are useful and perhaps necessary, and are somewhat analogous to feature-selection in machine learning tasks.

As a future task we would like to work on multiple-documents summarization evaluation.

REFERENCES

[1] Wang, Ruiyi, et al. "Microblog summarization using Paragraph Vector and semantic structure." Computer Speech & Language57 (2019): 1-19.

[2] Radev, D. R., HOVY, E., AND MCKEOWN, K. Introduction to the special issue on summarization. Computational Linguistics, 28:4 (2002), 399–408.

[3] Hassel, Martin. "Evaluation of automatic text summarization." Licentiate Thesis, Stockholm, Sweden (2004): 1-75.

[4] Hahn, Udo, and Inderjeet Mani. "The challenges of automatic summarization." Computer 33.11 (2000): 29-36.

[5] Kallimani, Jagadish S., K. G. Srinivasa, and B. Eswara Reddy. "Summarizing newspaper articles: experiments with ontology-based, customized, extractive text summary and word scoring." Cybernetics and Information Technologies 12.2 (2012): 34-50.

[6] Spark Jones, K.—Galliers, J.R.: Evaluating Natural Language Processing Systems: An

[7] Analysis and Review. In Lecture Notes in Artificial Intelligence, No. 1083, Springer, 1995.

[8] Lin, Chin-Yew, et al. "An information-theoretic approach to automatic evaluation of summaries." *Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics. Association for Computational Linguistics*, 2006.

[9] Lin, Chin-Yew, and Eduard Hovy. "Manual and automatic evaluation of summaries." Proceedings of the ACL-02 Workshop on Automatic Summarization-Volume 4. Association for Computational Linguistics, 2002.

[10] Lin, Chin-Yew. "Rouge: A package for automatic evaluation of summaries." Text Summarization Branches Out (2004).

[11] Luhn, Hans Peter. "The automatic creation of literature abstracts." IBM Journal of research and development 2.2 (1958): 159-165.

[12] Das, D., AND MARTINS, A. F. A survey on automatic text summarization. Tech. rep., Language Technologies Institute, Carnegie Mellon University, November 2007.

[13] GOLDSTEIN, J., Ed. Proceedings of the ACL-05 Workshop: Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization (2005).

[14] Sarkar, Kamal. "Automatic single document text summarization using key concepts in documents." Journal of information processing systems 9.4 (2013): 602-620.

[15] Dumias, S. T. Latent semantic indexing (lsi): Trec-3 report. *In Overview of the Third Text Retrieval Conference* (1995), pp. 219–230.

[16] Geiss, Johanna. Latent semantic sentence clustering for multi-document summarization. No. UCAM-CL-TR-802. University of Cambridge, Computer Laboratory, 2011.

[17] Bird, Steven, Ewan Klein, and Edward Loper. Natural language processing with Python: analyzing text with the natural language toolkit. " O'Reilly Media, Inc.", 2009.

[18] Ozsoy, Makbule Gulcin, Ferda Nur Alpaslan, and Ilyas Cicekli. "Text summarization using latent semantic analysis." Journal of Information Science 37.4 (2011): 405-417.

[19] Steinberger, Josef, and Karel Jezek. "Using latent semantic analysis in text summarization and summary evaluation." Proc. ISIM'04. 2004.

[20] Lin, Chin-Yew. "Rouge: A package for automatic evaluation of summaries." Text Summarization Branches Out: Proceedings of the ACL-04 Workshop. 2004.

[21] Miller, G. A. Wordnet: a lexical database for English. Commun. ACM 38, 11 (1995), 39–41.

[22] Mitchell, Tom Michael. The discipline of machine learning. Vol. 9. Pittsburgh, PA: Carnegie Mellon University, School of

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Computer Science, Machine Learning Department, 2006.

[23] Joachims, T. A probabilistic analysis of the rocchio algorithm with tfidf for text categorization. In Proceedings of International *Conference on Machine Learning (ICML)* (1997), pp. 143–151.

[24] Dias, Gaël, Elsa Alves, and José Gabriel Pereira Lopes. "Topic segmentation algorithms for text summarization and passage retrieval: an exhaustive evaluation." AAAI. Vol. 7. 2007.

[25] Angheluta, Roxana, Rik De Busser, and Marie-Francine Moens. "The use of topic segmentation for automatic summarization." *Proceedings of the ACL-2002 post-conference workshop on automatic summarization*. 2002.

[26] Kamble, Satish, et al. "Survey on Summarization Techniques and Existing Work." International Journal of Applied Engineering Research 12.1: 2017.

[27] Deerwester, Scott, et al. "Indexing by latent semantic analysis." Journal of the American society for information science 41.6 (1990): 391-407.

[28] Khorsi, Ahmed, Hadda Cherroun, and Didier Schwab. "A Two-Level Plagiarism Detection System for Arabic Documents." Cybernetics and Information Technologies 20 (2018).

[29] Foltz, Peter W., Walter Kintsch, and Thomas K. Landauer. "The measurement of textual coherence with latent semantic analysis." Discourse processes 25.2-3 (1998): 285-307.

[30] Gayathri, P., and N. Jaisankar. "Towards an Efficient Approach for Automatic Medical Document Summarization." Cybernetics and Information Technologies 15.4 (2015): 78-91.

[31] Kolluru, B., AND GOTOH, Y. On the subjectivity of human authored short summaries. In Goldstein [6], p. 916.