The 4th International Conference on Electrical Engineering and Informatics (ICEEI 2013)

CRF Based Feature Extraction Applied for Supervised Automatic Text Summarization

Nowshath K. Batcha[a],*, Normaziah A. Aziz[b], Sharil I. Shafie[a]

[a]*School of Computer Science and Information Technology,Taylors University,47500, Malaysia*
[b]*Kulliyah of Information and Communication Technology,International Islamic University Malaysia,50728,Malaysia*

Abstract

Feature extraction is the promising issue to be addressed in algebraic based Automatic Text Summarization (ATS) methods. The most vital role of any ATS is the identification of most important sentences from the given text. This is possible only when the correct features of the sentences are identified properly. Hence this paper proposes a Conditional Random Field (CRF) based ATS which can identify and extract the correct features which is the main issue that exists with the Non-negative Matrix Factorization (NMF) based ATS. This work proposes a trainable supervised method. Result clearly indicates that the newly proposed approach can identify and segment the sentences based on features more accurately than the existing method addressed.

1. Introduction

Information Overload (IO) refers to the state where a person's efficiency in using information in their work is hampered by too much of information available to them [1]. As the digital world has too much of information it becomes necessary for new measures which can give abstract or shortened summary version of the original text [2-4]. Automatic Text Summarization (ATS) is an automated process of detection and selection of relevant parts of

* Corresponding author. Tel.: +603-56295629.
  E-mail address: nowshath.kadharbatcha@taylors.edu.my

documents to be included in a summary. As information is overwhelmed on any topics as digital documents this process summarizes the texts automatically by condensation of texts preserving the meaning of the source documents. Actually it is described as the process of reducing the size of a text while preserving its information content by Luhn in 1958 [5]. Mani et.al explained in their work that text summarization filters the most important information from sources to produce an abridged version of it [6]. The primary goal of automatic text summarization is being source document as input, extract content from it, and output the most important content to the user in a condensed form. Jezek and Steinberger claims this task is basically a data reduction process and its necessity are enormously increasing due to easy availability of information on the World Wide Web [7]. Actually the growth of online textual databases and growth of the internet are the main reasons to make automatic text summarization as an active research area.

The remainder of this paper is organized as follows: Section 2 describes related work of ATS which gives us a clear background of ATS and Non-negative matrix factorization (NMF); Section 3 describes the existing problem with NMF in feature extraction when applied for ATS and Section 4 presents the proposed method and finally, in Section 5 we conclude the paper with limitations and future directions for enhancements.

2. Related work

Automatic Text Summarization also had gained considerable researchers attention due to its significance in Information Retrieval and Natural Language Processing. Many approaches to automatic text summarization have been proposed to identify and extract the most important information in a document. A Study from Neto, Freitas and Kaestner reveals that there are many orthogonal views of characterization of summarizations but the most often cited are Extracts and Abstracts [8]. Extracts are summaries comprising the important sentences extracted from the original document and presented as summary. These summaries even though suffer inconsistencies, they can conveniently represent an approximate content of the text for relevant judgment Abstracts, on the other hand, are concise summary of the central subject in the document and contain word sequences not present in the original. Up to now it is too hard task for computer research to solve creation of abstracts successfully [8]. Jing emphasizes that automating the text summarization by human is a complicated task to be modelled mathematically or logically [9]. Several studies reports that, people prefer extractive summaries instead of abstractive summaries. Further studies reveal that automatic text summarization can be type supervised or non-supervised. Supervised type [10, 11, 12, 13] characteristically takes assistance from human-made summaries or extracts to find features or parameters of summarization algorithms, while unsupervised type finds the relevant parameters without the assistance of human-made summaries to automatic text summarizers.

Many approaches or methodologies for automatic text summarization have been proposed to identify and extract the most important information in a document [14]. Lexical approaches for term-based text summarization had gained most researchers attention. One approach identified the important terms using lexical chains [15][4]. Another approach used co-reference or anaphora [16-19]. Further, it had also been demonstrated that significant performance improvement can be achieved when both approaches are mixed [20]. In [20], Singular Value Decomposition (SVD) and the output of an automatic anaphoric resolver were combined to get better performance. Even though several approaches for ATS are in practice recently algebraic reduction methods have grabbed the researcher's attention due to its dimensionality reduction factor. SVD and Non negative Matrix Factorization (NMF) are the most cited methods.

The basic idea behind SVD is taking a high dimensional, highly variable set of data points and reducing it to a lower dimensional space that exposes the substructure of the original data more clearly and orders it from most variation to the least. Ju-Hong et al demonstrated that since SVD can have negative values, (Non-negative Matrix Factorization) NMF can be used to select more meaningful sentences [21]. The same was also emphasized by various works that the major drawback of SVD is the possibility of negative value occurrences in the decomposed matrices. These negative components became an important pitfall of SVD in its application to text summarization.

2.1. Non negative Matrix Factorization

Recently Non-negative Matrix Factorization (NMF) has seized a finite attention in the field of information retrieval. Especially in the facet of ATS, it has gained considerable improvement in minimizing the gap between the summary produced by the man and machine. Ju-Hong et al proposed that as an alternative to SVD, NMF can be used to address this issue. This work also had revealed that the semantic feature vectors obtained by using the NMF are sparser than those obtained using SVD. Thus, the scope of the meaning of the semantic features obtained by using the NMF is clearer and narrower than that obtained by using SVD [21]. As proposed by Lee and Seung in 1999, NMF decomposes a given matrix into two non-negative matrices W and H [22] which is formally represented in (1).

$$A_{n \, x \, m} \approx W_{n \, x \, k} H_{k \, x \, m} \tag{1}$$

The primary problem in mining information from the texts is finding the semantic feature representation in thetexts.NMF does this by decomposing the term by sentence matrix which is generated from the source text to be summarized as shown in (1). Here $A_{n \, x \, m}$is the term by sentence matrix and$W_{n \, x \, k}$is called as non-negative semantic feature matrix (NSFM) whereas $H_{k \, x \, m}$ is called as non-negative sematic variable matrix (NSVM).The features or most important information that the sentences need to convey is captured in $W_{n \, x \, k}$ and the weights of each feature available in each sentence is captured in$H_{k \, x \, m}$. As proposed by Lee and Seung NMF decomposes a given matrix into two non-negative matrices W and H [22]. NMF algorithm keeps updating W and H until the object function converges under the predefined threshold or exceeds the number of repetition by using (2) and (3) given below:

$$H_{\alpha\mu} \leftarrow H_{\alpha\mu} \frac{(W^T A)_{\alpha\mu}}{(W^T \, WH)_{\alpha\mu}} \tag{2}$$

$$W_{i\alpha} \leftarrow W_{i\alpha} \frac{(AH^T)_{i\alpha}}{(WHH^T)_{i\alpha}} \tag{3}$$

NMF uses a repetitive or iterative technique to tune the values of W and H so that their combination approaches A. This procedure terminates when specified number of iterations is reached or when the approximation error converges. The NMF decomposition is non-unique; the values in W and H depends mainly on the NMF algorithm chosen, initial values of W and H chosen and the error measure used to check the convergence. For the error convergence mostly two cost functions are used as proposed in [22]. They are squared error (Frobenius norm) and an extension of the Kullback-Leibler divergence to positive matrices. Using the Frobenius norm for matrices, the convergence cost function is shown in (4). The other multiplicative method is shown in (2) and (3).

$$\min_{W,H} \|A - WH\|_F^2 \tag{4}$$

2.2. Role of NMF in Automatic Text Summarization

Fig. 1 explains the generic document summarization as proposed in [21]. It clearly shows the role of NMF in ATS. Actually the document to be summarized is pre-processed. Pre-processing involves three techniques which are tokenization, stop word removal and stemming. Tokenization is the process of breaking a stream of text up into words, phrases, symbols, or other meaningful elements called tokens. The list of tokens becomes input for further processing.Stop word removal is elimination of dead words like a, an and the as these words have no semantic meaning.And stemming is the process of identification of word stems and separating them from prefixes. Stemming is needed to obtain the root word. After this process, the document is converted it to term by document matrix which is very high in dimension. Hence for dimensionality reduction NMF is applied to this term by document matrix. It results in further two matrices as shown in (1).

After this process using the H matrix which carries semantic values of the sentences are scored using Generic Relevance Score (GRS) shown in (5) and (6) for inclusion in summary. Finally, the sentences with high values of generic relevance are selected for the summary to be included.
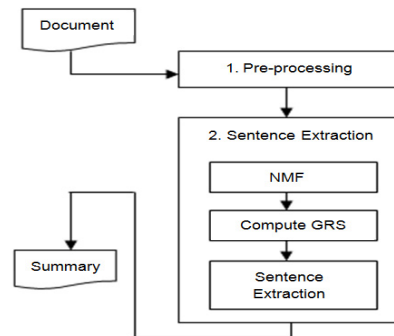

Fig. 1 Generic summarization method using NMF [21].

$$\sum_{i=1}^{r}(H_{ij} \cdot weight(H_{i*})) \qquad (5)$$

$$Weight (H_{i*}) = \frac{\sum_{q=1}^{n} H_{iq}}{\sum_{p=1}^{r} \sum_{q=1}^{n} H_{pq}} \qquad (6)$$

2.3. Example representation of a sentence using NMF

We illustrate an example of sentence representation using the NMF method by giving an example. The non-negative matrix A is a term-by-sentence matrix which is obtained after pre-processing, is further approximated to two matrices named as W and H. Each sentence can be represented as a linear combination of semantic features and semantic variable values which is shown in Fig 2. Fig 3(a) represents the distribution of feature wise semantic values among the sentences and Fig 3(b) represents the sentences with distribution of various semantic values in it. Scoring of sentences is done using (5) and (6) as proposed in [21].
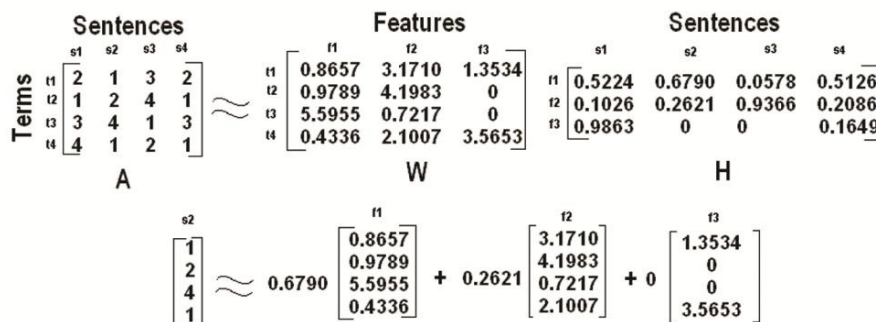

Fig.2. Example of a sentence representation using NMF.

Here H matrix gives the semantic values of each sentences corresponding to features f1, f2 and f3. When NMF applied to ATS, each sentence can be scored based on semantic weight values. H matrix weightage can be represented graphically as shown in Fig 4.
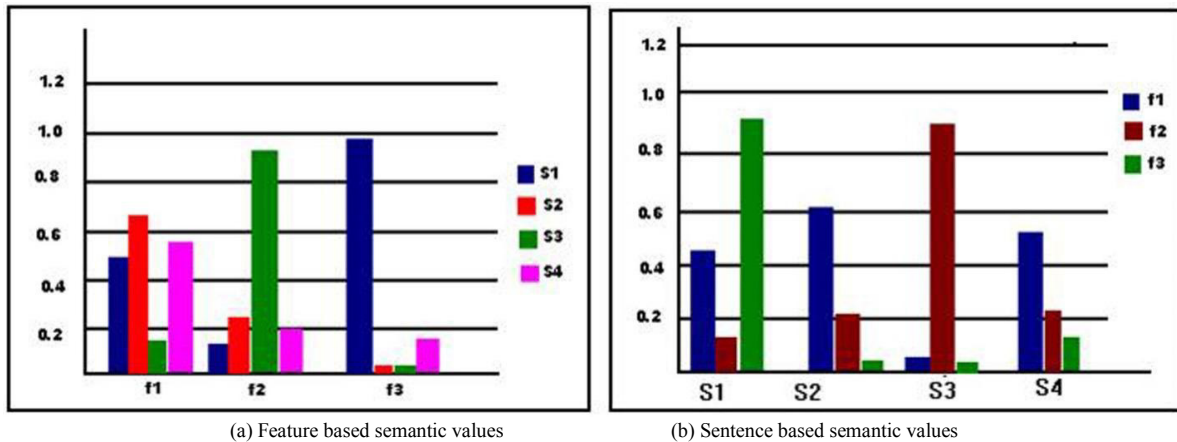
| (a) Feature based semantic values | (b) Sentence based semantic values |

Fig. 3. Graphical representation of semantic values of H matrix
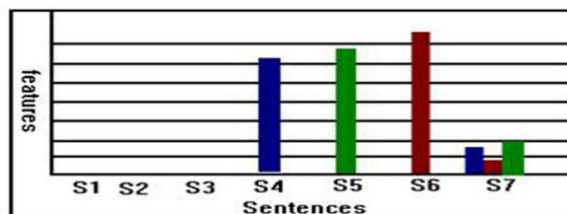

Fig. 4. Graphical representation of semantic values of H matrix-existing method

3. Existing problem of NMF in automatic text Summarization

To illustrate the problem when NMF used for ATS, we have chosen a set of sentences in an online news magazine shown in Table 1.The terms-by-sentences matrix A obtained by pre-processing the set of sentences in Table 1 which is composed of 95 terms (n) and 7 sentences (m). Further r is chosen to be 3 meaning extracting the information in matrix$A_{95 \times 7}$into decomposed matrices of $W_{95 \times 3}$and$H_{3 \times 7}$. The semantic feature vectors obtained in W is shown in Fig 5.

Table 1: Sample sentences

Sentence No	Sentences
S1	US Golf Association announced that courses will hose 36-hole sectional qualifying rounds for the US open golf championship.
S2	Hollywood legend Steven Spielberg is to make a television mini-series about Napoleon based on a screenplay by Stanley Kubrick.
S3	CIMB Equities Research is maintaining its Neutral outlook with WCT and Mudajaya as its top picks.
S4	The action movie star turned politician will become group executive editor for the magazines Flex and Muscle & SYM Fitness.
S5	University Teknologi Malaysia and the Iskandar are in discussion to develop a dedicated halal industrial park in Iskandar Malaysia.

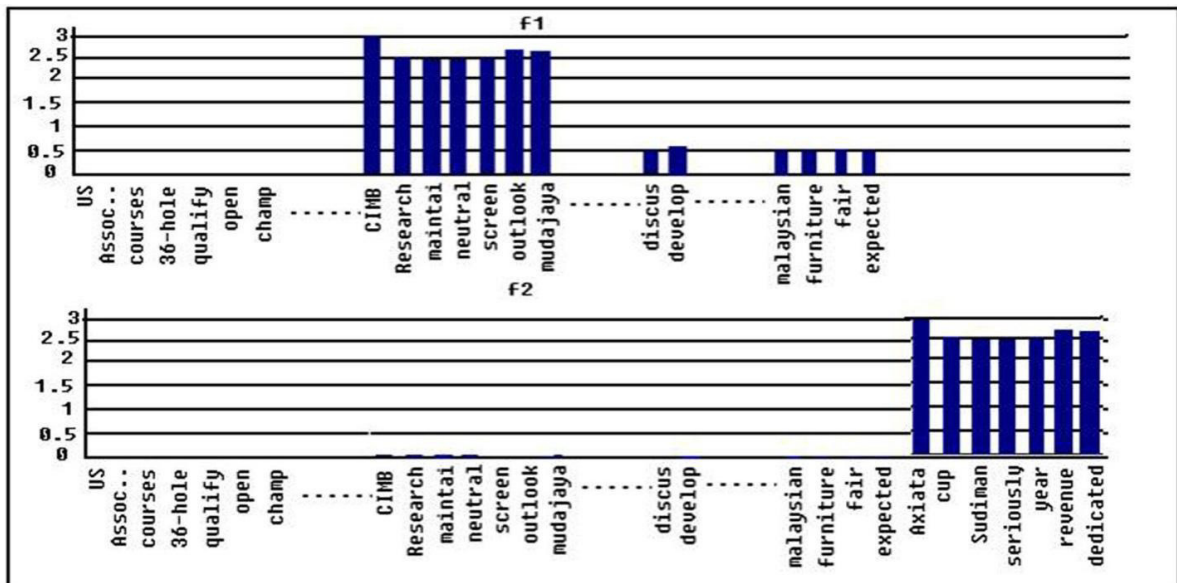| S6 | BAM general manager reported that it will take Axiata Cup Seriously since this year being Surdiman Cup Year. |
| S7 | The Malaysian International Furniture Fair is expected to hit at least US$871.5mil in revenue this year. |

Fig 5. Semantic Feature vectors –W matrix of the existing method

Observing the Fig 5, we cannot grasp what feature vectors f1 and f2 are representing to. Further when we look at the sentences taken for example shown in Table 1, we can notice that sentences S1 and S6 are about sports, S2 and S4 are about entertainment and S3, S5 and S7 are about business. The three graphs respectively for f1and f2 indicate that the semantic features from NMF consist of very few terms that have important meanings. Even though it can be taken as an advantage as stated in [21] but when it comes to representation of f1 and f2 we cannot conclude that it is inclined to sports, entertainment or business, since it has the scattered mixture of terms from all the three topics namely sports, entertainment and business. The same applies to f3. This would lead to linguistic noise which could lead to degraded performance in the automatically generated summaries.

Further let us look at the Sematic Value matrix H generated using NMF shown in Fig 4. The values of S1, S2 and S3 columns directs that sentences S1 , S2 and S3 did not contribute to any of the features of f1, f2 and f3 and in reality it is not the case. This happens because NMF is basically an approximation algorithm which is focused to approximate $W_{n \times k}$ and $H_{k \times m}$ matrix close to $A_{n \times m}$. Here only the values of W and H matrices are tuned based on the algorithm selected and initial values seeded for W and H matrices. Strictly no knowledge about the features available in the term by sentence matrix is considered. In fact it is not needed for the NMF algorithms as its intended for approximation and not designed for feature extraction.

Its primary focus is approximation and the focus on preserving the features of the source matrix in $W_{n \times k}$ is of secondary importance and this leads to linguistic noises in the representation. This will obviously degrade the quality of summary generated by ATS when using NMF and concretely this is the problem which this work is addressing to. Hence we propose a new method which could consider the addressed problem specific to domains using a training set to construct the initial seeds for W and H matrices. These initial seeds of W and H lead to converge the optimal solution preserving the knowledge obtained from the training set. Further the semantic feature and semantic value matrices generated are reflected based on the semantic content of the term by sentence matrix A.

4. Proposed domain specific summarization method

The proposed domain specific approach is clearly explained in Fig 6. Actually this approach makes use of any already available or built corpus to construct the training sets. Any document can be viewed as a collection of sentences that can be grouped into several segments. Each segment can have representative sentences which can be informative about that segment. Since all the sentences belong to the same domain most of them can have predefined features with mostly likely occurring patterns. These patterns can be the training sets that can be used for segmentation. The proposed approach first and foremost builds the training set identifying patterns that can be found in the domain specific corpus. Once identified and given a sequence of test sentences, Conditional Random Fields (CRF) segmentation will label the sentences so that the likelihood of the label sequence of the training set given the whole test sentence sequence is maximized. The intension of the proposed approach here is classifying the sentences based on the patterns to segments. Once grouped or classified we would be able to identify the terms contributing to that segment or topics. Those segments or topics apparently become the feature vectors of the W matrix. In this method CRF is used as a tool to model the sequence labelling problem [23]. Hence the proposed method can extract the sentence based features more meaningful than the existing NMF method in the $W_{n \times k}$ matrix.
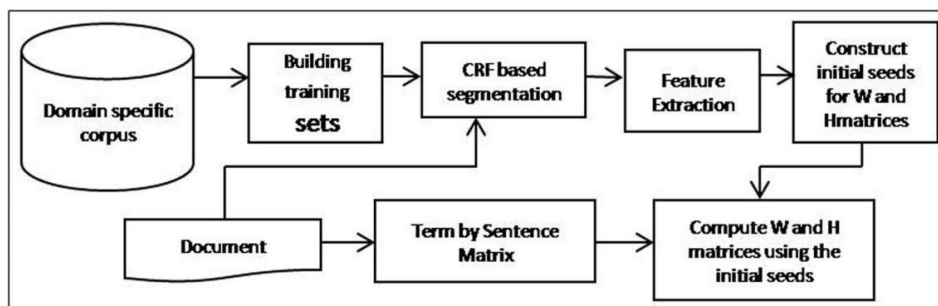

Figure 6: Proposed method

The input to the CRF segmentation task consists of a sequence of words or called as tokens. For example US Golf Association announced that courses will hose 36-hole sectional qualifying rounds for the US open golf championship is the sequence of 18 tokens: [US, Golf, Association, announced, that, courses will, hose, 36-hole, sectional, qualifying, rounds, for, the, US, open, golf, championship]. Segmenting a token sequence $[T_1 \dots T_n]$ with CRF amounts to assigning a label sequence $[L_1 \dots L_n]$ where each L_i is assigned to token T_i from the predefined training set. So for any given token sequence *X,* CRF computes the label sequence *Y* with the highest conditional probability $P(y|x)$ as the best label sequence. For an exposition how CRF segments token sequences using training labels, let us take the sentence S1 for example given in Table 1, we have constructed the training patterns using the POS tagging as shown in the Table 2.

Here you can notice that each column shows the patterns constructed where each line in a column can be called as template [23].Each template defines the semantics among the columns. For example, 1st column is 'word', second column is 'POS tag' and the last column represents a true answer tag which is going to be trained by CRF.

Table: 2 Sample patterns

Pattern1	Pattern2	Pattern3	. . . . .	Pattern n
US NNP SBN	Courses NNS SBN			
open JJ SBA	will MD SBM	US NNP SBN		UTM NNP BBN
golf NN SIN	hose NN SIN	golf NN SIN	. . . . .	and CC BBC
championship NN SIN	36-hole NN SIN	association NN SIN		ISKANDAR NNP BIN
will MD SBM	sectional JJ SBA	stated VBN SIV		are VBP BBV
have VB SBV	………	…….		………
………				

You can notice that in the answer tag we have prefixed the topic or segment label as S for sports, B for Business and E for Entertainment. The rest of the characters follow the IOB2 format [24].So when a test statement is given without the answer tag CRF tool can give answer tag which can indicate the segment or topic it belongs to. Once sentences are classified we can construct the feature sets with the terms contributing to them and construct the initial seeds of $W_{n \times k}$ and $H_{k \times m}$ matrices. The procedure used to construct the $W_{n \times k}$ and $H_{k \times n}$ matrix is shown in Fig 7(a) and Fig 7(b).

TermArray[] is an array of all the terms in the document and Feature_Set[] is an array of all the Features available in the document. Here W_{ij} is populated with initial seeds using the weight of number of occurrences of TermArray[i]th term in the Feature_Set[N]th feature divided by the count of total number of terms which contributes to Feature_Set[N]th feature. Here we group the terms available in the document based on segments or topics and each sement becomes the column vector of the $W_{n \times k}$ matrix.

(a) $W_{n \times k}$ matrix   (b) $H_{k \times m}$ matrix
Fig. 7. Procedures for initial seed construction

Hence at this point we avoid the linguistic noise available with the previous NMF based approach because we can clearly identify the features of $W_{n \times k}$ matrix and each term related to that segment or topic. SentenceArray[] is an array of all the sentences in the document and Segment_set[] is an array of all the segments or topics available in the document. Here H_{ij} is filled with initial seeds using the weight of number of occurrences of SentenceArray [i] th sentence in the Segment_Set[N]th segment divided by the count of total number of sentences which contributes to Segment_Set[N]th segment.

At this point the linguistic noise available in $H_{k \times m}$ matrix shown in Fig 4 can be avoided.The main purpose of populating the $W_{n \times k}$ and $H_{k \times m}$ matrices with good seeds are that even though NMF has many advantages in applying to information retrieval, it comes with a caveat: it must be properly initialized and the initialization selected is crucial to getting good solutions [25]. All NMF algorithms are iterative and sensitive to initializations. Some NMF algorithms require both W and H matrices to be properly initialized [26-29]. To sum up a good initialization can improve the speed and accuracy of the algorithms, as it can produce faster convergence to an improved local minimum. Once W and H matrices are fed with initial seeds then the proposed method computes the W and H matrices to extract the semantic variables and semantic values. Experimentation with this approach shows better results can be achieved than the existing NMF method which is explained in the next section.

5. Experimentation and Results

We used the sample sentences shown in the Table 1 for experiments. The proposed method can extract the sentence based features more meaningful than the existing NMF method. Further study also shows that better representation of the sentences can be achieved using the proposed method.First we experimented the feature vectors obtained using the existing method and the proposed method. Graphical representation of the feature vectors obtained using the existing method is shown in Fig 5. Observe that features f1, f2 and f3 are spread over terms through which we cannot conclude what each of the features are representing like sports, entertainment or business. But in contrast, observe the feature vectors resulted using the proposed approach in Fig 8, you can notice that terms are finely grouped appropriately to represent like f1 for sports, f2 for entertainment and f3 for business.
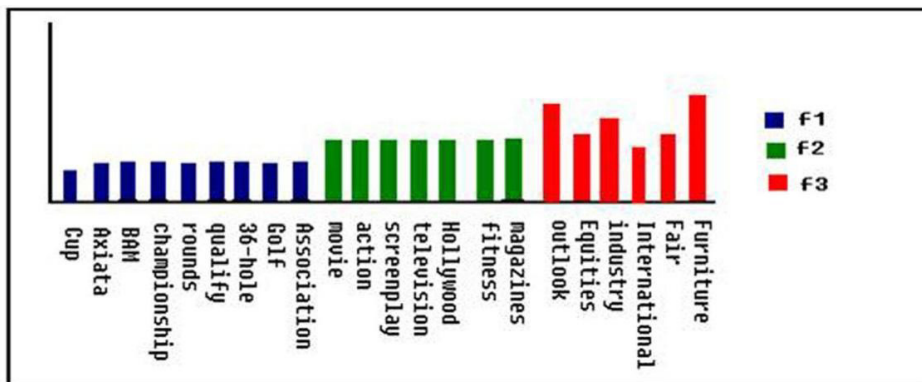

Fig. 8. Semantic Feature Vectors of W matrix- Proposed Method

Further the H matrix obtained using the proposed method is also shown in Fig 9. When you compare the results obtained for the H matrix graphically shown in Fig 4 and Fig 9 you can understand that the proposed method can represent the semantic weights in the sentences better than the existing method.
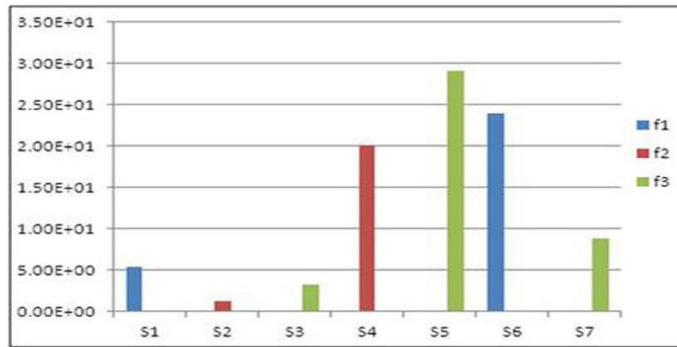
Fig. 9.: Semantic Values of Sentence in H matrix of the proposed method

A comparative chart of both the methods is shown for better understanding in Fig 10. Observe that proposed method can clearly shows that sentences labeled S3,S7,S5 contributes with semantic values for the feature f3 (business) , sentences S4 and S2 contributes to f2 (entertainment) and finally S1 and S6 to f1 (sports) respectively. Overall the proposed method can extract the semantic features and values in W and H matrices more accurate than the existing method is illustrated clearly with this example. Hence in the process of ATS when W and H matrices are computed as proposed in this study would give better results in the summary generated automatically.

6. Conclusion and Future Work

The results obtained clearly shows the proper initialization of W and H matrices plays a vital role in the results obtained. Further we have considered simple generic sentence structures for the experimentation which is one of the limitations. Hence our future leaps aimed at considering complex sentence patterns for the study. It is also aimed at applying the proposed study for more specific domains like law, medical and news summarization.
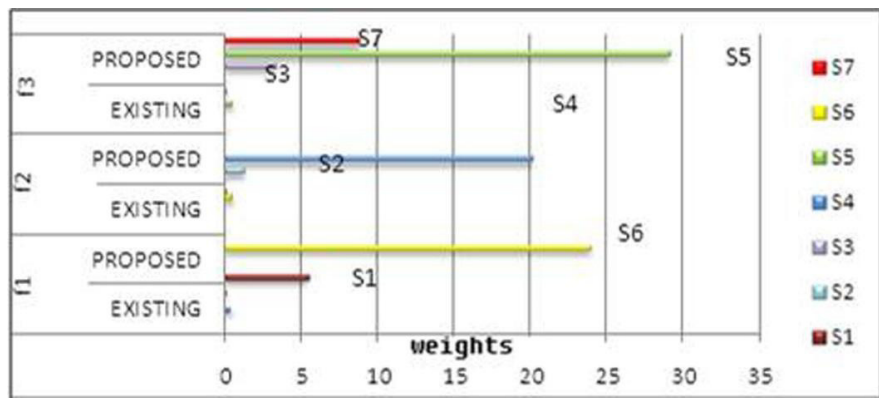

Fig. 10. Comparison of semantic values in H matrix between the existing and proposed methods

References

[1] Yang C.C, Chen H,and Hong Visualization of large category map for Internet browsing, Decision Support Systems, 2003: 35 (1): 89–102.

[2] Li Cheng, Automatic Text Summarization based on Rhetorical Structure Theory, Computer Application and System Modeling (ICCASM), 2010 International Conference on, 2010: 13: V13-595,V13-598.

[3] Ou, Shiyan, Christopher KhooS.G. and Dion Hoe-LianGoh. Automatic Text Summarization in Digital Libraries. Handbook of Research on Digital Libraries: Design, Development, and Impact. IGI Global, 2009:159-172.

[4] Gong Y, and Liu X, Generic text summarization using relevance measure and latent semantic analysis .Proceeding of SIGIR, 2001.

[5]LuhnH. P, The automatic creation of literature abstracts. IBM J. Res. Dev. 2, 2, 1958: 159-165.

[6] Mani, Inderjeet, and Mark T. Maybury, eds. Advances in automatic text summarization. MIT press, 1999.

[7] Josef Steinberger, Massimo Poesio, Mijail A. Kabadjov, KarelJežek, Two uses of anaphora resolution in summarization, Information Processing & Management, 2007: 43(6) :1663-1680.

[8] Neto, J., Freitas, A., &Kaestner, C. Automatic Text Summarization Using a Machine Learning Approach: Proceedings of the 16th Brazilian Symposium on Artificial Intelligence: Advances in Artificial Intelligence, London, UK, 2002.

[9] Jing H and McKeownK.R., Cut and Paste based text Summarization , Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference, Morgan Kaufmann Publishers Inc, 2000.

[10] Terano T, LiuH., and. ChenA. L. P, Knowledge Discovery and Data Mining, 1805 in Lecture Notes in Artificial Intelligence, pages 454–457. Springer, 2000.

[11] Kupiec, J, Pedersen J, and Chen F,.A Trainable Document Summarizer. In Proceedings of the Fourteenth Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval, Seattle,USA, 1995: 68–73.

[12] BergerA. and Mittal V. O, Query-relevant summarization using faqs. In Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics, 2000: 294–301.

[13] Marcu D, The automated construction of large-scale corpora for summarization research. In Proceedings of the 22nd International ACM/SIGIR Conference on Research and Development in Informational Retrieval, 1999:137–144.

[14] Hovy E, and Lin C, Automated Text Summarization and the SUMMARIST system, in Proc. of a ACL/EACL Workshop held at Baltimore, Madrid Spain, 1998: 197-214.

[15] BarzilayR. and ElhadadM., Using Lexical Chains for Text Summarization, in Proc. of ACL/EACL Workshop on Intelligent Scalable Text Summarization, Madrid, Spain,199: 10-17.

[16] BaldwinB. and Morton T.S, Dynamic Coreference-based Summarization, in Proc. of the Third Coference on Emprical Methods in Natural Language Processing, Granada, Spain, 1999: 1-6.

[17] Boguraev B and Kennedy C, Salience based Content Characterization of Text Documents, in Mani I and Maybury M., eds., Proc. of ACL/EACL '97 Workshop on Intelligent Scalable Text Summarization, Madrid, Spain, 1997: 2-9.

[18] Bergler S, Witte R, Khalife M, LiZ. and Rudzicz F, Using Knowledge-poor Coreference for Text Summarization, in Proc. of HLT/NAACL (Human Language Technologies/North American Chapter of the Association for Computational Linguistics) Workshop on Document Understanding Conference, Edmonton,2003.

[19] StuckardtR., Coference based Summarization and Question Answering: a Case for High Precision Anaphor Resolution, in International Symposium on Reference Resolution, Venice, Italy, 2003.

[20] Steinberger J, Massimo P, Sanchez-Graillet O Improving the LSA based Summarization with Anaphora Resolution, in Proc,of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, Vancouver, 2005: 1-8.

[21] Lee, Ju-Hong and Park, Sun and Ahn, Chan-Min and Kim, Daeho,Automatic Generic Document Summarization based on Nonnegative Matrix Factorization, Information Processing and Management: an International Journal, 2009: 45: 20-34.

[22] Lee D.D. and Seung H.S, Learning the parts of objects by nonnegative matrix factorization , Nature, 1999: 401: 788-791.

[23]Lafferty, John, Andrew McCallum, and Fernando CN Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. 2001.

[24] Aramaki, Eiji, Takeshi Imai, Kengo Miyo, and Kazuhiko Ohe. Automatic deidentification by using sentence features and label consistency. In i2b2 Workshop on Challenges in Natural Language Processing for Clinical Data, 2006: 10-11.

[25] Langville, Amy N., et al. Initializations for the nonnegative matrix factorization. Proceedings of the Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2006.

[26] Patrik O. Hoyer. Non-negative sparse coding. In Neural Networks for Signal Processing XII (Proc.IEEE Workshop on Neural Networks for Signal Processing), 2002: 557–565.

[27] Patrik O. Hoyer. Non-negative matrix factorization with sparseness constraints. Journal of Machine Learning Research, 2004: 5:1457–1469.

[28] Lee D and SeungH..Algorithms for Non-Negative Matrix Factorization. Advances in Neural Information Processing Systems, 2001: 13:556–562.

[29] Paul Pauca V, Jon Piper, and Robert J. Plemmons.Nonnegative matrix factorization for spectral data analysis. 2005.