Finding Summary of Text Using Neural Networks and Rhetorical Structure Theory

Sarda A. T.[1], Kulkarni A. R.[2]

[1]Research Scholar, Computer Science & Engineering, Walchand Institute of Technology, Solapur, India

[2]Assistant Professor, Computer Science & Engineering, Walchand Institute of Technology, Solapur, India

Abstract: *A new technique for summarization is presented here for summarizing articles known as finding summary of text using neural network and rhetorical structure theory. A neural network is trained to learn the relevant characteristics of sentences by using back propagation technique to train the neural network which will be used in the summary of the article. After training neural network is then modified to feature fusion and pruning the relevant characteristics apparent in summary sentences. Finally, the modified neural network is used to summarize articles and combining it with the rhetorical structure theory to form final summary of an article.*

Keywords: backpropagated neural networks, rhetorical structure theory, text summarization, concession

1. Introduction

Automatic text summarization is the technique, where a computer find summary for given text document. A text document is given as input to the computer a summarized text document is returned as output, which is a non redundant extract from the original text. The technique has its ideas in the 60's and has been developed during 30 years, but today with the Internet and the World Wide Web the Automatic text summarization technique has become more important.

With the explosion of the WWW and the abundance of text material available on the Internet, text summarization has become an important and timely tool for assisting and interpreting text information. The Internet provides more information than is usually needed. Therefore, a twofold problem is encountered: searching for relevant document through an massive number of articles available, and absorbing a large amount of related information. Summarization is a useful to selecting related articles, and for extracting the important points of each articles. Some articles such as academic papers have accompanying abstracts, which make them easier to decipher their important points. However, sport articles have no such accompanying summaries, and their titles are often not sufficient to convey their key points. That's why, a summarization tool for articles would be very useful, since for a given topic or event, there are a big number of available articles from the various web portals and newspapers. Because sport articles have a highly structured document form, important ideas can be obtained from the text simply by selecting sentences based on their attributes and locations in the article. [3]

We propose a machine learning approach that uses neural networks to produce summaries of articles. A neural network is trained for articles. The neural network is then modified, through comparing & combining feature, to produce highly ranked sentences for summary of the article. Through feature fusion, the network discovers the importance (and unimportance) of various features used to determine the summary-worthiness of each sentence. [3]

2. Neural Network

Neural Networks are made up of the layers. Layers are made up of a number of 'nodes' which are interconnected & contain an 'Activation function'. Patterns are presented to the network via the 'input layer', which communicates to 'hidden layers' where the actual processing is done via a system of weighted 'connections'.

It is a Multi-layer feed forward or back propagation in architecture. In neural network architecture the information flows from input layer to output layer. It consists of one input, one or more hidden layer and one output layer. From input layer inputs are sent into units then weighted output from these units are taken as in the next layer that is hidden layer, weighted output of this layer is sent as input in the next hidden layer and so on. Until output of last hidden layers is send to output layer. Output layer gives the result which is predicted output.
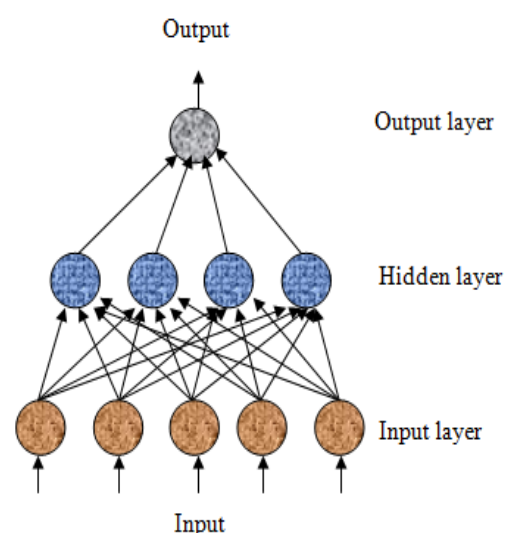

Figure 1: Neural Network

Paper ID: NOV163987

2260

3. Features

Each article is converted into a list of sentences. Each sentence is represented as a vector $[f_1,...,f_8]$, made up of 8 features. Given as bellow,

Table 1: Features

F_1	Paragraph follows title.
F_2	Paragraph location in document.
F_3	Sentence location in paragraph
F_4	First sentence in paragraph
F_5	Sentence length
F_6	Number of thematic words in the sentence
F_7	Number of title words in the sentence
F_8	Numerical data feature

Feature f_1 Paragraph follows title, which finds location of paragraph here first paragraph which follows title feature f_2 Paragraph location in document, which finds location of paragraph among all paragraph present in document. feature f_3 Sentence location in paragraph, which finds sentence location among all sentences from paragraph and decides rank for sentences as per their position. Feature f_4 first sentence in paragraph which decide sentence score and rank by its position in paragraph in this case first sentence in paragraph. Feature f_5, sentence length, is useful for finding out long and short sentences such as dateline and names commonly found in different articles. We also anticipate that short sentences are unlikely to be included in summaries. [3] Feature f_6, the number of thematic words, which point out the number of thematic words in the sentence, relative to the maximum possible words according to the theme of article. Feature f_7 Number of title words in the sentences, which indicates the number of title words in the sentence, relative to the maximum possible. [3] Feature f_8 Numerical data feature is used find numerical data in sentences to find more feasible sentence for summary.

4. Rhetorical Structure Theory

RST addresses text organization by means of connection that grasp between parts of text. It explains coherence by postulating a hierarchical, connected structure of texts. Rhetorical relations or coherence relations or discourse relations are paratactic (coordinate) or hypotactic (subordinate) relations that hold across more than one text spans. It is widely accepted that notion of coherence is through text connection like this. Rhetorical Structure Theory using rhetorical relations provide a methodical way for an analyst to analyse the text. An analysis is usually constructed by reading the text & building a tree using the relations. The example given below is a title and summary, the original text, broke down into units having numbers, is:
1. The Perception of Apparent Motion
2. When the motion of an intermittently seen object is ambiguous
3. the visual system resolves confusion
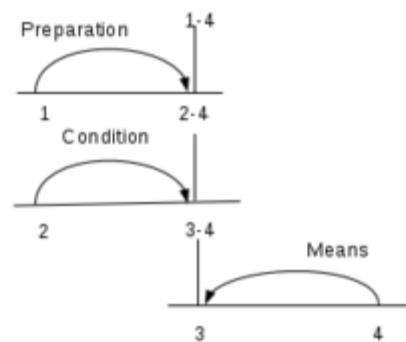4. by applying some tricks that reflect a bulletin knowledge of properties of the physical world

Figure 2: Rhetorical Relations

In the figure2 number 1,2,3,4 displaying the correspond units as explained above. 4th unit and 3rd unit forming a relation Means. 4th unit is the important part of this relation. So it is known as nucleus of the relation and 3rd unit is known as satellite of the relation. Similarly 2nd unit to 3rd and 4th unit is forming relation Condition. spans may be composed of two or more units.[16]

5. Methodology

In this system user gives article as input document. Then document is converted into sentences. Each sentence is represented in a vector form created by features. After that actual summarization process starts.

There are some phases in process of neural network training, feature combining & feature selection and sentence selection. The 1st phase involves neural network training to identify the type of sentences that should be inserted in the summary.

The 2nd phase, feature combining which also called as feature fusion, feature selecting which is also called as feature pruning by applying both to the neural network which give away the hidden layer unit activations into discrete values with frequencies. This phase finalise features that must included in the summary sentences by combining the features and finding fashion in the summary sentences.
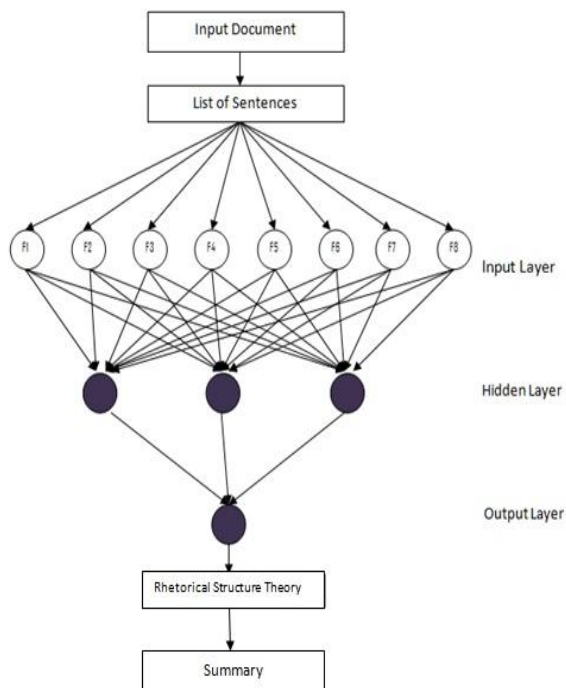
Figure 3: Methodology

The 3rd phase, sentence selection, uses the modified network to find the text & to select only the highly ranked sentences in the summary. This phase controls the selection of the summary sentences in a way of their importance and rank & produces basic summary used to give as input to the rhetorical structure theory which finally produces the final summary.

5.1 Neural Network Training

The 1st step of the process includes training the network to learn the types of sentences that should be involved in the summary. This is completed by training the neural network with sentences in several paragraphs where each sentence is identified as to whether the sentence should be taken in the summary or should not taken in summary. This is done by a human reader. The neural network learns the patterns inherent in sentences that should be taken in to the summary & those that should not be included. It can find the patterns and approximate the inherent function of any data to accurate it up to the mark, as long as there are not any contradictions in the data set. Our neural network consists of 8 input layer neurons, 5 hidden layer neurons, and 1 output layer neuron. We use a Gaussian method where the energy function is a combination of error and penalty function. The aim of training is to find for the global minima of the energy function. The addition of the penalty function drives the associated weights of un-necessary connections to very small values while strengthening the rest of the connections. Therefore, without affecting the performance of the network we can prune unnecessary connections & neurons.

Here in our network for training purpose we use 0.9 as learning rate, 0.00001 is maximum error value & 100000 is max iterations. We are using back propagation neural network with 8 input nodes, one hidden layer with 5 nodes & one node at output layer. So generally how it works let's see,

Here's how we calculate the total net input for any node in hidden layer is as below.

$$\mathbf{NET\ H} = \left(\sum w_a * I_a \right) + b_1 * 1$$

EQ 1. Net input for any node in hidden layer.

Here, H= Any node in hidden layer,
W= Weight of incoming node,
I= Node from input layer,
a= Number of node from input layer starting from 1,
b_1= Bias value from node of input layer to node of hidden layer.

By using above formula we can find out net input to any node of hidden layer. Now we need to find out output from any node of hidden layer, which can be done by usin formula given below.

$$\mathbf{OUT\ H} = \frac{1}{1 + e^{-NET\ H}}$$

EQ 2. Output from any node in hidden layer.

By using above formula we can find out output from hidden layer, using the output from the hidden layer neurons as inputs. NET O is derived by formula given below,

$$\mathbf{NET\ O} = \left(\sum W_a * OUT\ H \right) + b_2 * 1$$

EQ 3. Input for node in output layer.

Here, Net O=input for node of output layer,
b2=Bias value for node of hidden layer to node of output layer.

After finding all outputs and inputs for neural network now it's time to calculate total error by using formula given below. We can now calculate the error for each output neuron using the squared error function and sum them to get the total error,

$$TE = \frac{\sum (target - output)^2}{2}$$

EQ 4. Total Error.

Here, TE=Total Error,
Target=Ideal Output,
Output=Actual Output.

Our goal with back propagation is to update each of the weights in the network so that they cause the actual output to be closer the target output, thereby minimizing the error for each output neuron and the network as a whole.

$$\frac{d\ TE}{dw} = d_0 OUT\ H$$

EQ 5. Partial derivative of TE with respect to w.
Where,

$$d_0 = -(target_0 - OUT_0) * OUT_0 (1 - OUT_0)$$

EQ 6. formula for finding value for d_0

Finally, we've updated all of our weights! But not necessary that new output are close to our target but repeating back propagation process several times we might be closer to our target output, Means with very less errors as we defined 0.00001 max error limit .

5.2 Sentence Selection

Once the network has been trained, pruned, and generalized, selection process include a process to find sentences in paragraph and determine whether each sentence should be included in the summary or not. This step is accomplished by providing control parameters for the radius and frequency of hidden layer activation clusters to select highly ranked sentences from neural network. The sentence ranking is inversely proportional to cluster radius & directly proportional to cluster frequency. Only sentences that satisfy the required cluster boundary and frequency are selected as high-scored summary sentences.

5.3 Rhetorical Structure Theory

After finding high ranked summary sentences by neural network we feed articles to rhetorical structure to find the discourse structure from that and find rhetorical relation in sentences which may help in finding better summary sentences, which further might be used to form better summary.

As we know that there are many rhetorical relations available such as Summary, Restatement, Preparation, Concession .Summary, satellite presents a restatement of the content of nucleus, that is shorter in bulk, where reader recognizes satellite as a shorter restatement of nucleus. Restatement, satellite restates nucleus, where satellite and nucleus are of comparable bulk; nucleus is more central to writer's purposes than satellite is, where readers recognizes satellite as a restatement of nucleus. Preparation, satellite precedes nucleus in the text; satellite tends to make readers more ready, interested or oriented for reading nucleus. Concession, writers acknowledges a potential or apparent incompatibility between nucleus and satellite, recognizing the compatibility between nucleus and satellite increases reader's positive regard for nucleus. We are using all these rhetorical relations to sort out sentences from article and combine these sentences with previously sorted sentences by neural network. And form final summary as result.

6. Literature Review

In the previous research, different techniques were presented for producing summary of any text or articles.

Khosrow Kaikhah presented "Text Summarization Using Neural Networks", this technique is used to the selection of features as well as the selection of summary sentences by the human reader from the training paragraphs plays an important role in the performance of the network. The network is trained according to the style of the human reader and to which sentences the human reader deems to be important in a paragraph. This, in fact, is an advantage our approach provides. Individual readers can train the neural network according to their own style. In addition, the

selected features can be modified to reflect the reader's needs and requirement.[3]

M. KarthiKeyan & K. G. Srinivasagan, represented "Multi-Document and Multi- Lingual Summarization using Neural Networks", this technique is used to generate multi-document summarization, and describes the details of each step. The performance of the text summarization process depends predominantly on the style of the human reader. The selections of features as well as the selection of summary sentences by the human reader from the training paragraphs play an important role in the performance of the network. The neural network is trained according to the style of the human reader and to which sentences the human reader deems to be important in paragraph Individual readers can train the neural network according to their own styles. In addition, the selected features can be modified to reflect the reader's needs and requirements. To generate precise summarization, more in-depth understanding of the sentence (paragraph) is required.[6]

W.T. Chuang and J. Yang represented "Extracting sentence segments for text summarization: a machine learning approach" this technique is used to design of automatic text summarizer. It will reduce the pain of people suffer reading huge amounts of data by offering them a cosine summary for each document. They developed an automatic text summarizer based on sentence segment extraction. It generates a summary based on the rules derived from any superwised machine learning algorithm.[7]

Nicolaos B. Karayiannisrepresented "A Methodology for Constructing Fuzzy Algorithms for Learning Vector Quantization", in this technique he presented a new methodology for constructing FALVQ algorithms, which exploits the fact that the competition between the winning and nonwinning prototypes during the learning process is regulated by the interference functions.[8]

7. Experimental Setup & Results

The system was developed using Java platform and using neuroph for creating and training neural network here we use 100 sports documents to train neural network and 50 sport documents to test output of neural network. Here we use human generated summaries also known as reference summaries to compare with our summaries and also used Copernicus Summarizer to find and compare summaries with our summaries.

To compare summaries we used ROUGE package known as Recall-Oriented Understudy for Gisting Evaluation. It includes measures to automatically determine the quality of a summary by comparing it to ideal summaries created by humans. The measures count the number of overlapping units such as n-gram, word sequences, and word pairs between the computer-generated summary to be evaluated . the ideal summaries created by humans. Here in ROUGE we use 3 formulas for finding summary results are Recall, Precision & f-measure. Recall means ratios of no. of matched n-grams between test summary and reference summary by no. of n-grams of reference summary. Precision means ratios of no. of matched n-grams between test

summary and reference summary by no. of n-grams of test summary. F-Measure means ratios of recall and precision.

We divided 50 sports documents in 5 sets let's see their results and comparison with human generated summaries all tables are give as below.

Here, Copernicus=Copernic Summerizer,
NN=Neural Network,
NN with RST=Neural Network with Rhetorical Structure Theory.

Table 2: Results of set 1

Doc 1-10	Copernicus	NN	NN with RST
Recall	0.267222206	0.4515246	0.572236184
Precision	0.419834229	0.6624083	0.584466804
F-Measure	0.312244052	0.5071557	0.553320603

Table 2 shows results in means of recall, precision & f-measure for set 1 containing average of 1-10 documents in sport category where recall, precision & f-measure of our algorithm NN with RST are better than Copernicus as well as NN. Graph for same is shown below
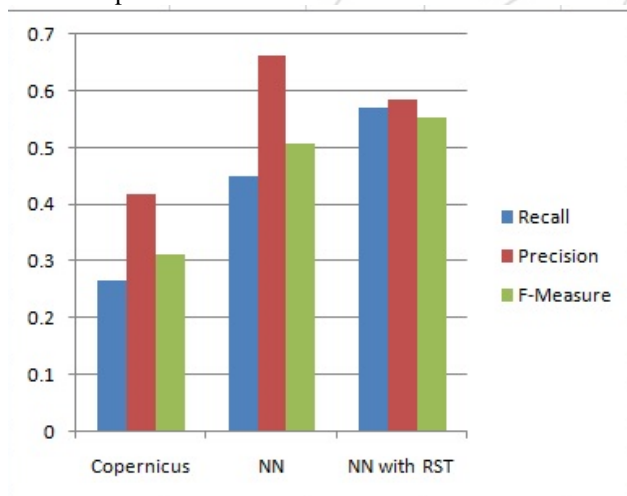

Figure 4: Comparison graph for set 1

Table 3: Results of set 2

Doc 11-20	Copernicus	NN	NN with RST
Recall	0.365968691	0.458919	0.506108432
Precision	0.499888586	0.713259	0.685844844
F-Measur	0.411199979	0.541166	0.559515496

Table 3 shows results in means of recall, precision & f-measure for set 2 containing average of 11-20 documents in sport category where recall, precision & f-measure of our algorithm NN with RST are better than Copernicus as well as NN. Graph for same is shown below
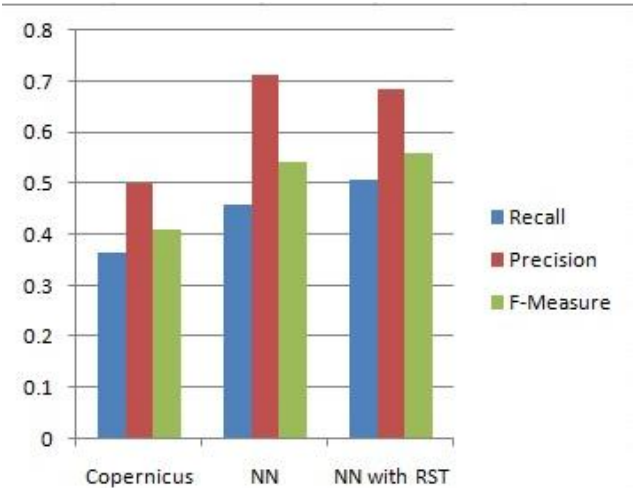

Figure 5: Comparison graph for set 2

And so on we compare remaining sets as above and at last we found average of all data set as given below,

Table 4: Results of all documents

Doc 1-50	Copernicus	NN	NN with RST
Recall	0.35748123	0.481378	0.555821953
Precision	0.54263098	0.72804	0.657180017
F-Measur	0.413323259	0.553907	0.576289937

Table 4 shows results in means of recall, precision & f-measure for all sets containing average of 1-50 documents in sport category where recall, precision & f-measure of our algorithm NN with RST are better than Copernicus as well as NN. Graph for same is shown below
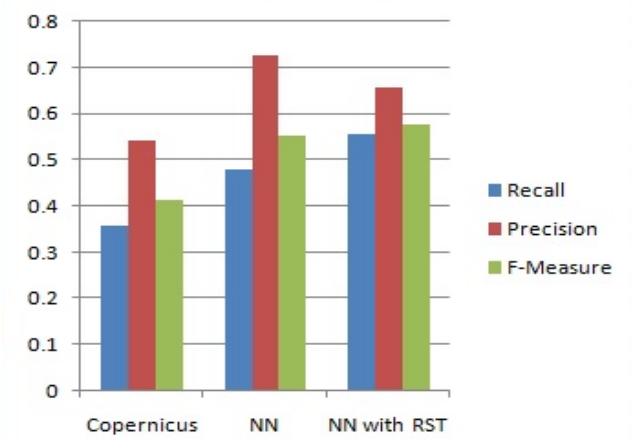

Figure 6: Comparison graph for alldocuments

8. Conclusion

Now a day's most of the people prefer to read summary of any document instead of reading whole document because the summary includes core part of the document. The selection of features & the selection of summary sentences to form better summary using neural network. By comparing to previous research done by KhosrowKaikhah[3] we include new feature Numerical data feature, which will help to select highly ranked summary sentences. Also we

modified neural network from feed forward neur withal network to backpropagation neural network which gives better training with minimum error to neural network & also generate better summary than previous. Also we modified our program by addig Rhetorical Structure Theory provides a combination of features that useful in several kinds of discourse studies & also provide some features to form better summary than previous. As we can see in our results f-measure of our algorithm that is NN with RST is 16.29% more efficient than copernicus & 2.23% moer efficient than NN.

References

[1] M S Patil, M S Bewoor, S H Patil," Survey on Extractive Text SummarizationApproaches",NCI2TM: 2014.

[2] Md. MajharulHaque, SuraiyaPervin, and Zerina Begum," Literature Review of Automatic Multiple Documents Text Summarization", International Journal of Innovation and Applied Studies ISSN 2028-9324 Vol. 3 No. 1 May 2013, pp. 121-129 2013.

[3] KhosrowKaikhah "Text Summarization Using Neural Networks", Department of Faculty Publications-Computer Science, Texas State University, eCommons,2004.

[4] Vishal Gupta & Gurpreet Singh Lehal, "A Survey of Text Summarization Extractive Techniques", Journal Of Emerging Technologies In Web Intelligence, Vol. 2, No. 3, August 2010.

[5] J. Kupiec, J. Pederson and F. Chen, "A Trainable Document Summarizer", Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Seattle, Washington, pp. 68-73, 1995.

[6] M.KarthiKeyan&K.G.Srinivasagan," Multi-Document and Multi-Lingual Summarization using Neural Networks" International Conference on Recent Trends in Computational Methods, Communication and Controls (ICON3C 2012) Proceedings published in International Journal of Computer Applications (IJCA)

[7] W.T. Chuang and J. Yang, "Extracting sentence segments for text summarization: a machine learning approach", Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Athens, Greece, pp. 152-159, 2000.

[8] Nicolaos B. Karayiannis," A Methodology for Constructing Fuzzy Algorithms for Learning Vector Quantization", IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 8, 1997

[9] Guangbing Yang, Dunwei Wen, Kinshuk, Nian-Shing Chen and ErkkiSutinen," Personalized Text Content Summarizer for Mobile Learning: An Automatic Text Summarization System with Relevance Based Language Model", IEEE Fourth International Conference on Technology for Education, 2012

[10] Julian Kupiec, Jan Pedersen and Francine Chen, "A Trainable Document Summarizer" Xerox Palo Alto Research Center 3333 Coyote Hill Road, Palo Alto, CA 94304.

[11] Ms.PallaviD.Patil, Prof.N.J.Kulkarni, "Text Summarization Using Fuzzy Logic" International Journal of Innovative Research in Advanced Engineering (IJIRAE) Volume 1 Issue 3 (May 2014) SPECIAL ISSUE

[12] Sandra A. Thompson, Wlliam C. Mann," Rhetorical Structure Theory: A Frarnework for the Analysis of Texts", IPM Papers in Pragmatics I, No.1 , 79-lO5. (1987)

[13] Simon H. Corston-Oliver, "Identifying the linguistic Correlates of Rhetorical Relations", Microsoft research one Microsoft way, Redmond WA 98052-6399 USA.

[14] Eva Forsbom "Rhetorical Structure Theory in Natural Language Generation", Uppsala University and GSLT GSLT: Natural Language Generation Teacher: Hercules Dalianis Spring 2005.

[15] Nick Nicholas," Parameters for Rhetorical Structure Theory Ontology", University of Melbourne

[16] Mr. Sarda A.T., Mrs. Kulkarni A.R.," Text Summarization using Neural Networks and Rhetorical Structure Theory", International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 6, June 2015