Grenze International Journal of Engineering and Technology, Jan Issue

Automatic Text Summarization using Wu-Palmer Measure and Graph based Sentence Selection

Rishav Rakshit[1], Chandralika Chakraborty[2], Udit Kr. Chakraborty[3] and Bhairab Sarma[4]
[1]Microsoft Corporation, Hyderabad, India
Email: rishav.rakshit@microsoft.com
[2-3] Sikkim Manipal Institute of Technology, (SMU), Majhitar, India
Email: {chandralika.c, udit.c }@smit.smu.edu.in
[4]University of Science & Technology Meghalaya, Baridua, India
Email: sarmabhairab@gmail.com

***Abstract*—In this paper, an extraction based text summarization methodology is proposed that aims to overcome issues faced with traditional extraction based summarizers using a combination of testing semantic similarity using the Wu-Palmer Measure and a graph-based sentence selection scheme for single document summarizations. The proposed approach is not only independent of context, syntax and language but also promotes shorter summaries via its sentence selection scheme. This method has been tested using the standard Reuters-21578 collection against widely used TextRank based summarizers and has shown promising results for single-document summarizations.**

***Index Terms*— Automatic, Text Summarization, Wu Palmer, Graph, extraction.**

I. INTRODUCTION

The present decade is witnessing amelioration of connectivity. The massive reach of the internet and its rate of growth has increased web-based content manifold. This has necessitated the development of text summarizers which can be used to retrieve relevant information out of large texts. Text summarizers capable of delivering succinct content, retaining the originality of the source document find large scale applications in the fields of finance [1], journalism [2], medicine [3], internet search engines and even generic research [4].

A summary, according to Radev et.al [5] is defined as "a text that is produced from one or more texts, that conveys important information in the original text(s), and that is no longer than half of the original text(s) and usually, significantly less than that." Automatic text summarization therefore can be taken as the process of automatically producing the summary from a given text.

There exists two approaches to automatic text summarization, namely extractive summarization and abstractive summarization. Extractive summaries consist of important sentences extracted from the parent text and copied verbatim into the summary. Abstractive summaries, in contrast try to generate fresh sentences conveying the most critical information from the text to be summarized. Traditional extractive summarizers do not guarantee a coherent narration [6] but can represent an approximation of the content of the text with relative ease.

The current paper proposes a method of extractive summarization of single documents. The method proposed,

Grenze ID: 01.GIJET.8.1.521

uses a graph based method to find the similarity between sentences using the Wu-Palmer method [7] and a relevant lexical database, WordNet[8]. The proposed summarizer is unsupervised, not specific to content, can be adapted to multiple languages, maintain consistent narration and adheres to the percentage of summarization specified by the user.

II. Basic Approach to Extractive Text Summarization

All text summarizers perform three standard tasks which are independent of each other. These tasks are [9]:
i) building an intermediate representation of the input text
ii) scoring the sentences of the intermediate representation
iii) selection of sentences to build the summary
Building intermediate representation is popularly done using either topic representation or indicator representation [10]. Topic representation approaches use frequency of topic words to identify topics, while the indicator driven approaches use indicators such as sentence length, word position of phrase detection.
In this regard, methods such as term frequency[11], TF/IDF [11] and resemblance to title have been shown to be most effective [12] for sentence selection, but they have their own limitations as well. Word frequency is highly dependent on the document format and writing style while TF/IDF is suited to mainly multi-document applications. Resemblance to title also suffers through a similar fate as word frequency.
The proposed method uses lexical similarity between sentences, which determines the coherence between two sentences of a text depending on their semantics rather than syntax – thus the quality of the final summary is based on the 'meaning' of the text rather than 'how' the text has been written.

III. Theoretical Background to Sentence Scoring and Selection Strategy

To determine the relative scoring between 2 sentences or nodes, the proposed approach uses a combination of the Wu-Palmer measure and a lexical database such as WordNet, which itself is represented as linked synonym sets. The summarizer itself can be extended to languages other than English by using the relevant WordNet-like lexical system such as the IndoWordNet[13].

A. The Wu-Palmer Similarity Measure

The knowledge-based similarity measure proposed by Wu and Palmer computes the similarity between concepts through edge counting, parent concepts and distance from the hierarchy root. The relatedness of two words is calculated by considering the depths of the two synsets in the WordNet taxonomies, along with the depth of the Least Common Subsumer (LCS). The generated score when 2 synsets of 2 words are compared is calculated as shown in expression (i):

$$\textbf{score (S) = 2 * depth (lcs) / (depth (s1) + depth (s2))} \qquad \textbf{(1)}$$

This means that $0 < \text{score} <= 1$. The score can never be zero because the depth of the LCS is never zero (the depth of the root of a taxonomy is one). The score is one if the two input concepts are the same.
Comparison between similarity measures carried out by Lin [14] suggests that Wu-Palmer measure is simple to calculate, and has high performance without losing precision. The selection of the Wu-Palmer method to determine sentence similarity is therefore justified.
In the proposed approach, if N1 and N2 as shown in Fig 2, represents the Nodes containing two sentences of the input text. Then each word of N1 is compared to each word of N2 and the Wu-Palmer scores are generated and summed to produce a cumulative score (CS). This cumulative score is assigned as the weight of the edge between N1 and N2.
Suppose s1 and s2, as shown in Figure 1 are two synsets on w1 and w2 such that w1 belongs to N1 and w2 belongs to N2. R is the root node of the synsets and the CS is the common ancestor between W1 and W2. The diagrammatic representation for Wu-Palmer score calculation is shown in Figure 1.

B. Graph Based Sentence Selection Scheme

The graph based sentence selection scheme determines the sentences of the original text that get selected into the final summary that is to be generated. It relies on calculating the cumulative scores of all the edges between a starting node and all the other nodes linked to that node. Once the relative score of all the edges have been determined, the node belonging to the edge containing the highest cumulative score is selected. This node is said to have the strongest relation to the starting node.

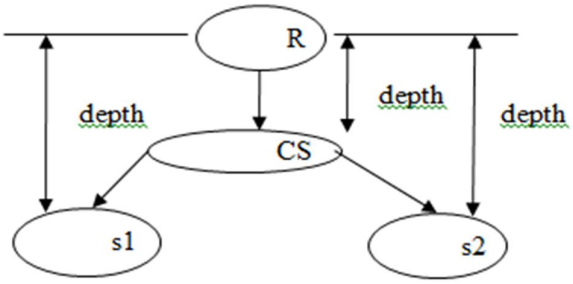

Figure 1: Concept Hierarchy in WordNet

The selected node now becomes the starting node and the cumulative scoring and selection process is applied to all other nodes linked to the current starting node barring the nodes that already have been selected. This process is repeated until the percentage of summarization as specified by the user is achieved.

The graph based sentence selection scheme is considered to be effective in its reduction of computation time as the Wu-Palmer similarity measure need not be applied to all the nodes of the graph but only to those nodes which are deemed to have the greatest semantic similarity to the most recently selected node, as shown in Figure 2.

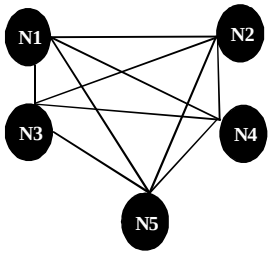

Figure 2: Node Connectivity

TABLE I. INITIAL MATRIX REPRESENTATION OF SENTENCE SELECTION SCHEME

	N1	N2	N3	N4N5	
N1	X	3.89	4.56	5.56	3.20
N2	X	X			
N3	X		X		
N4	X			X	
N5	X				X

TABLE II. SELECTED SENTENCES

Selected Nodes	N1	N2	N3	N4	N5
	1	0	0	1	0

TABLE III. MATRIX REPRESENTATION OF SENTENCE SELECTION SCHEME AFTER 1ST SENTENCE SELECTION

	N1	N2	N3	N4	N5
N1	X	3.89	4.56	5.56	3.20
N2	X	X			
N3	X		X		
N4	X	2.43	3.88	X	3.10
N5	X				X

IV. Proposed Approach and Algorithm

The proposed approach involves the Wu-Palmer similarity measure and the graph based sentence selection working in tandem with each other. A single document text is taken as input which is summarized by a given percentage in an unsupervised fashion. The complete process, detailed herein below comprises of four distinct steps; namely preprocessing, mapping, sentence scoring and selection and output generation.

A. Preprocessing

In the preprocessing phase, the input text is sanitized, removing all punctuation marks and abbreviations. The stop words are kept relatively intact, as although they do not contribute significantly to lexical similarity, such words are useful for establishing syntactical coherence between sentences, which is one of the objectives of this paper.

B. Mapping

The preprocessed text is now split into distinct sentences and each of these sentences are mapped to a graph. Each sentence is treated as a node of the graph, joined by an edge to every other node. The graph thus generated, is strongly connected in nature.

C. Sentence Scoring and Selection

The sentence scoring and selection is performed according to the following algorithm. It is to be noted that the starting node (**S**) is the first sentence of the input text, as set accordingly in 85% of all texts, the starting sentence is deemed to be the most important one[15]. It is also the first sentence to be selected into the summary by default.

*Algorithm Selection (**G, S, P**)*

Input – The Mapped Graph (**G**) containing N Nodes/ vertices V connected through edges,

where **N** : Number of Sentences,

S : Starting Node,

P : is the number of nodes to be selected as deemed by the desired percentage of summarization.

Output – A Set of Nodes (**M**) containing the selected sentences.

Step 1: Start
Step 2: M={M U S}
Step 3: While (N!=P) do
Step 3.1 For each edge E in G such that that E∈{S,V} do
Step 3.1.1: Extract all words of S and V, and compare each word of S with each word of V using Wu-Palmer similarity to derive a cumulative similarity score for S versus V.
Step 3.1.2: Assign this score as weight of E.
Step 3.2: End For
Step 3.3: Select U having maximum edge weight.
Step 3.4: M = {M ∪ V}
Step 3.5: Perform G = G – {S}
Step 3.6: Assign S→V
Step 4: End While
Step 5: Stop

D. Output Generation

In the output generation phase, the sentences that appear in M as nodes are displayed according to the order in which they appear in the original text. If continuation sentences such as those beginning with he, she, this, that, as, but etc. are present in M then they can be handled in 2 ways:

i) Either these sentences are completely removed from the summary[16].
ii) Or the sentence occurring immediately prior to this continuation sentence is included in the summary and subsequently the terminating sentence of the summary is removed.

Finally a post-processing phase is performed where the punctuations and short forms appearing in the initial text are restored. This can be performed in multiple methods such as maintaining a binary array to keep track of the sentences that have been selected from the original document and their positions in the said document as

depicted in Figure2 and Tables I, II and III. The selected sentences are provided as output in the order in which they appear in the input text.

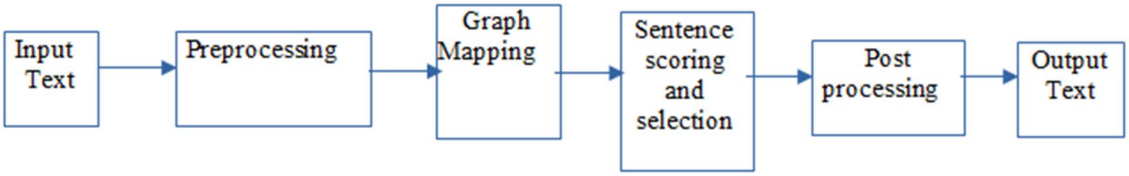

Figure 3. Summarization Process Flowchart

V. Evaluation and Results

Adhering to the proposed approach, an application is created in Python 2.7 programming language, implementing Wu-Palmer similarity using the English WordNet. The application is tested against one designed according to TextRank[17]. The evaluations are performed against 10 articles selected from Reuters- 21578, a standard text dataset used for testing natural language processing and informational retrieval algorithms.

A. Evaluation Methodology

In this methodology we have decided to reduce human intervention as much as possible and hence a more objective evaluation technique is utilized. Six articles of length of 11 to 35 sentences are selected from the Reuters – 21578 dataset and fed as input text to both summarizers. The generated summaries from the summarizers are then compared in the following manner:

i) The overlap sentences between both summaries are removed from the evaluation. All these overlaps are considered as 'correct' selections.

ii) Each non-overlap sentence of both summaries are taken and a fair expert selects the sentences that are most relevant to the context. The selected sentences are considered as 'correct' for their respective summarizers and the remaining sentences are deemed as 'wrong'. The judging is done on the basis of similarity with context, similarity to the title of the text and finally, preservation of continuity.

iii) Subsequently the Precision (P), Recall (R) and F-Measure (F) for each summarizer is calculated as:

$$Precision\ (P) = correct\ /\ (wrong+correct) \qquad (2)$$

Since both summarizers are summarizing up to 50% of the input text, the Recall (R) and F-Measure (F) are also the same as Precision (P) . It is also to be noted that the Precision (P) that is calculated for each summarizer is relative to each other and is not the absolute measure.

B. Test Results

Test	No. of Sentences	No. of Sentences in Summ.	No. of Overlaps (each)	No. of Non-Overlaps (each)	Correct (each)		Wrong (each)		Precision (P)	
					Text Rank –based	Proposed	Text Rank-based	Proposed	Text Rank-based	Proposed
Text 1	13	7	3	4	2	3	2	1	0.714	0.857
Text 2	27	13	10	3	2	3	1	0	0.923	1.000
Text 3	30	15	11	4	2	3	2	1	0.866	0.933
Text 4	26	13	10	3	2	1	1	2	0.923	0.846
Text 5	32	17	15	2	1	1	1	1	0.941	0.941
Text 6	21	10	6	4	2	3	2	1	0.800	0.900

The results obtained are very encouraging. In most of the tests the TextRank based summarizer and our application is neck and neck in performance. For the evaluation against Text 2 having the title 'LEADING INDUSTRIAL NATIONS TO MEET IN APRIL' , both summarizers performed at the same level. However the TextRank based summarizer lost out in the decision against one sentence which violated the continuity of the

summarized text. Similarly in Text 4, our summarizer suffers a similar fate due to lack of refinement of the post-processor.
The overall non-overlaps have ranged from 2 to 4 sentences which ranges the difference in judgment between the two summarizers to be at a minimum of 11.7% for Text 5. In the majority of cases the number of overlaps far exceeds the number of non-overlaps.

VI. Conclusion and Further Work

The proposed summarizer has shown good results against TextRank based summarizers. Its general independence of form, format and language of texts and processing of sentences allowed us to achieve all our five primary objectives.
Future prospects would include refining the post-processing phase, testing the approach against Machine-Learning based summarizers [17] and implementation of the approach for WordNets designed in other languages.

References

[1] Li, X., Xie, H., Song, Y., Li, Q, Zhu, S., Wang, F.L., "Does summarization help stock prediction? A news impact analysis." IEEE Intelligent Systems 30.3 (2015): 26-34.
[2] Gao, H., Barbier, G., Goolsby, R., "Harnessing the crowd-sourcing power of social media for disaster relief." IEEE Intelligent Systems 26.3 (2011): 10-14.
[3] Marcelo, F., Rindflesch, T. C. Kilicoglu, H., "Summarizing drug information in Medline citations." AMIA Annual Symposium Proceedings. Vol. 2006. American Medical Informatics Association, 2006.
[4] Qazvinian, V., Radev, D., R., "Scientific paper summarization using citation summary networks." Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1. Association for Computational Linguistics, 2008.
[5] Radev, D., R., Hovy, E., McKeown, K., 2002, “Introduction to the special issue on summarization”, Computational Linguistics, 28, 4 (2002), 399-408.
[6] Neto, J., L., Freitas, A., A., Kaestner, C., A., A., "Automatic text summarization using a machine learning approach.", Brazilian Symposium on Artificial Intelligence. Springer Berlin Heidelberg, 2002.
[7] Wu, Z., Palmer, M., “Verb semantics and lexical selection”. In Proceedings of the 32nd Annual meeting of the Associations for Computational Linguistics, pp 133-138. 1994.
[8] Miller, George A. "WordNet: a lexical database for English." Communications of the ACM 38.11 (1995): 39-41.
[9] Nenkova, A., McKeown, K.,” A survey of text summarization techniques.” Mining Text Data. Springer, 43–76, 2012.
[10] Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E.D., Gutierrez, J. B., Kochut, K., “Text Summarization Techniques: A Brief Survey” [Downloaded from: https://arxiv.org/abs/1707.02268]
[11] Murdock, Vanessa Graham. Aspects of sentence retrieval. Dissertation. University of Massachusetts, Amherst, 2006.
[12] Ferreira, R., Cabral L. deSouza., Lins, R. D., eSilva, E. P., Freitas, F., Cavalcanti, G. D. C., Lima, R., Simske, S., J., Favaro, L., "Assessing sentence scoring techniques for extractive text summarization." Expert Systems with Applications 40.14 (2013): 5755- 5764.
[13] Bhattacharyya, P. "IndoWordNet. The WordNet in Indian Languages.” Springer Singapore, 2017. 1-18.
[14] Lin, D., “An Information-Theoretic Definition of similarity”. In., Proceedings of the fifteenth International Conference on Machine Learning (ICML'98). Morgan-Kaufmann: Madison, WI, pp.296-304.1998.
[15] Das, D., Martins, A. F. T., "A survey on automatic text summarization." Literature Survey for the Language and Statistics II course at CMU 4 (2007): 192-195.
[16] Kedar, B., Sarma, A. D., Sarma, A. D. Loiwal, N., Mehta, V., Ramakrishnan, G., Bhattacharyya, P., "Generic Text Summarization Using WordNet." LREC. 2004.
[17] Summa NLP, 2014; GitHub: http://github.com/summanlp/textrank