ISSN (Print) : 0974-6846
Indian Journal of Science and Technology, *Vol 8(32), DOI: 10.17485/ijst/2015/v8i32/92118, November 2015* ISSN (Online) : 0974-5645

An Optimized Semantic Technique for Multi-Document Abstractive Summarization

Atif Khan*, Naomie Salim and Adekunle Isiaka obasa

Faculty of Computing, Universiti Teknologi Malaysia, 81310, Skudai, Johor, Malaysia;
atifkhan@icp.edu.pk, naomie@utm.my, iaobasa@yahoo.com

Abstract

Background/Objective: Multi-document summarization produces a concise summary from several online topically related documents. A major challenge in this domain is usually the information overlap in documents emanating from various sources. This paper introduces an optimized semantic technique for multi-document abstractive summarization. **Methods/Statistical Analysis:** Linguistic and semantic approaches are usually employed for abstractive summarization of multiple documents. Linguistic approaches lack semantic representation of source text while semantic approaches mostly rely on human experts to construct domain ontology and rules; which require immense time and effort. The technique in this paper utilizes the benefits of semantic role labeling, clustering and Particle Swarm Optimization (PSO) to rank predicate argument structures (semantic representation) in each cluster using optimized features. **Findings**: The summary quality is susceptible to the text features i.e., different features have varied importance towards summary generation. Therefore, optimal features weights obtained using PSO integrated in the semantic technique to rank semantic representation improved summarization results. The performance of the technique is evaluated against the benchmark summarization systems using pyramid evaluation measures (mean coverage score, precision and F-measure). A Paired-Samples T-test is carried out to validate the summarization results. **Applications/Improvements**: Experiment of this research is performed with DUC-2002, a benchmark data set for text summarization. Experimental results confirm that the proposed technique yields better results than other comparison summarization models in terms of mean coverage score and average F-measure.

Keywords: Language Generation, Multi-Document Abstractive Summarization, Particle Swarm Optimization, Semantic Similarity Measure, Semantic Role Labeling (SRL)

1. Introduction

Currently, online users are overwhelmed with the gigantic amount of text due to the fast expansion of information over the internet. Text summarization is demanded for such information overload. It is an important and timely tool for users to promptly understand the enormous amount of information. Multi-Document Summarization (MDS) chooses the most salient information from the text documents and produces a short summary that can suit user needs[1]. It facilitates online users to acquire information in efficient manner. It has gained more consideration in various research areas such as information retrieval, Natural Language Processing (NLP), and machine learning[2].

Approaches of text summarization are generally separated into two classes: Extractive and abstractive summarization. In extractive summarization, the salient sentences are chosen from the given text documents and grouped to create a summary. On the other side, abstractive summarization is a bit hard area; it requires the text to be represenated semantically, employs compression and natural language generation techniques. This type of summarization comprehends the source document text and generate a brief summary that typically contains compact sentences or may include some new sentences not appearing in the source document[3,4].

Major research have paid attention to multi-document extractive summarization utilizing sentence extraction techniques[5,6], techniques based on statistical analysis[7,8], different machine learning and discourse structures techniques[9,10]. Nowadays, the researchers are leaning

* *Author for correspondence*

towards abstractive summarization and has become a hot area due to the fact that a few research attempts have been made in this track. A peculiar demand for MDS is that information overlap usually exist in topically related documents. Therefore, appropriate summarization approaches are desirable to combine similar information emanating from numerous documents[11]. Previous abstractive summarization techniques rely on human experts to build domain ontology and rules; and then semantic representation of source document is built from them, which is a shortcoming of automated summarization system. However, we introduce an optimized semantic technique for multi-document abstractive summarization that will merge overlapping information from numerous related documents automatically, and employ language generation to form a brief abstractive summary. At first, the approach employs SRL to obtain Predicate Argument Structure (PAS), a semantic representation, from document text. Then, we utilize semantic similarity measure to analyze the text in order to group PASs across the text that are semantically similar. Finally, the PASs are ranked based on the different text features that are weighted and optimized by PSO; since summary quality is susceptible to the text features. Our contributions are given as follows:

- Introduce an optimized semnantic technique for Multi-Document Abstractive Summarization (MDAS).
- Investigate PSO to acquire optimal text features for summary generation.
- To evaluate the proposed technique with Pyramid measures on DUC 2002 shared tasks for MDAS.

The remainder of the paper is planned as follows: Section 2 presents the related literaure to this research. Section 3 gives detail of the proposed technique. Section 4 covers the results and discussion. Lastly, Section 5 demonstrates the conclusion of this research.

2. Related Work

Previous literature reveals that limited research efforts have been attempted for abstractive summarization. Researchers have utilized different methods to generate abstractive summaries. These methods are separated into two classes: Linguistic (Syntactic) and Semantic based approaches. The former approach uses syntactic parser to identify verbs and nouns in the text. So, verbs and nouns are used for syntactic representation of text and processed further to produce abstractive summary. However, semantic based approach uses semantic representation of document text for producing abstractive summary. Ontology and template based approaches represent text semantically and are widely sudied in the literature. Some researchers differentiated semantic and syntactic representation of sentence[12]. They stated that syntactic analysis is different at a greater extent from semantic represenation of sentences. Specifically, syntactic analysis does not describe who did what to whom.

The different linguistic based approaches[11,13–15] introduced for the abstractive summarization mainly depend on syntactic representation of text. The prevalent shortcoming of these approaches is that they do not represent the document text semantically.

A limited number of semantic approaches are also attempted for abstractive summarization, which are presented as follows. GISTEXTER is a MDS system discussed in[16], that generates abstractive summary from numerous newswire documents. It employs template based method to repersent topic of a document and relies on the outcome of Information Extraction (IE) systems. The major drawback of this system was that it required humans efforts to create extraction rules and linguistic patterns for the template slots.

A fuzzy ontology based summarization approach[17] was presented for Chinese news. It describes the domain knowledge in a better way by modeling uncertain information. The approach has a few limitations. First, the domain expert needs to define domain ontology and Chinese dictionary, which is time consuming. Secondly, the approach may not be applied to English news as it it has been introduced for Chinese news.

The methodology introduced in[18] employs abstraction shcemes to produce short abstractive summaries from clusters of news articles talking about a similar event. The abstraction scheme consists of information extraction module embedded with rules, heuristics for content selection, and a few generation patterns for producing sentence. However, the methodology required human expert knolwledge for making rules for Information Extraction (IE) and patterns for sentence generation. A framework presented by[19] represents multimodal document by a semantic model and produces abstractive summary from it. The semantic model is built from the knowledge represented by concepts of ontology. The concepts of semantic model are ranked using Information Density (ID) metric, and significant concepts are represented as sentences by making use of phrases saved for the concepts in the semantic model. However, the framework is soley dependent on domain expert to build domain ontology, and is not applicable to other domains.

The abstractive approach presented by[20] generates summary from Rich Semantic Graph (RSG) representaion of source document. RSG is contructed from ontology i.e., graph vertices are the occurrences of verb and noun classes in the ontology. However, this approach has some drawbacks. First, it relies on domain expert for constructing a domain ontology, which is limited to a particular domain; and in case the domain changes, the ontology will need to be rebuilt and therefore the semantic graph will be re-constructed. Secondly, this approach is applied only to single document and did not report any evaluations. A series of analysis studies is performed by[21] to compare human made summaries with system summaries based on semantic level of caseframes. However, these studies did not propose any summarization model. In[22] introduced an approach that exploits syntactic units such as noun/verb phrases to construct new sentences. The approach constructs facts and concepts from the document text and represent them by phrases. Finally, the approach employ an optimization algorithm to pick and combine the different phrases simultaneously. In[23] presented an approach that employed the idea of Basic Semantic Unit (BSU) to illustrate the meanings of an event. The approach captures semantic information of the text documents by defining a semantic linkage network which take into consideration BSUs. The sentences produced by the semantic link network constitute the structure of the summary.

The major shortcoming of the semantic based abstractive summarization approaches is that they generally rely on human experts to build domain ontology and rules; which is a limitation of an automated summarization system. Furthermore, these approaches need immense human effort and time and may not be pertinent to other domains. However, our work proposes an optimized semantic technique that will employ Semantic Role Labeling (SRL) to automatically construct semantic representation (PAS) from the source text. The PASs are further manipulated and the top ranked PASs are chosen for summary generation based on optimized features, obtained with PSO. Moreover, our approach has the potential to be applied to any domain. The next Section demonstrates our proposed technique.

3. Overview of Technique

The architecture of our proposed technique is depicted in Figure 1. At first step, we divide the document set (to be summarized) into sentences in a manner that each sentence is indexed by its associated document number followd by sentence location number.

Next, we employ semantic role parser[24] to obtain PAS from sentence collection in the document set. The semantic similarity matrix is constructed from pair wise similarities of PASs, which are determined using Jiang's similarity measure[25]. The time and location arguments of PASs are compared using edit distance algorithm.

Next, similarity matrix of PASs is given to agglomerative hierarchical clustering method to group PASs that are semantically similar. Section 3.3 will discuss this phase. The top ranked PASs from each cluster are chosen based on seleced features and their optimal weights are determined using particle swarm optimization (as described in Section 3.4.2).

Finally, we employ Simple NLG realisation engine[26] to produce sentences from the chosen PASs. The produced sentences will constitute the abstractive summary (as demonstraed in Section 3.5).

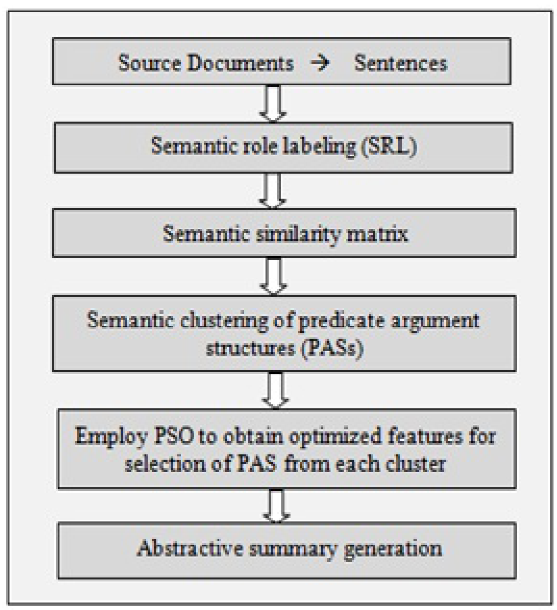

Figure 1. Proposed swarm semantic hybrid approach.

3.1 Semantic Role Labeling

The function this step is to obtain PAS from the senentce collection in the document set. Initially, the document set is divided into sentences in a manner that each sentence is indexed by its associated document number followed by sentence location number. As deep semantic analysis of text is required in case of abstractive summarization, so this work utilizes semantic role parser[24] to extract PAS from each sentence by properly labeling the word phrases in each sentence. The word phrases are called semantic arguments, which can be classified into two groups: Core arguments and adjunctive arguments[1]. This study considers the following core arguments: *A0* (subject), *A1* (object), *A2* (indirect object), and adjunctive arguments*:* *ArgM-LOC* (location), *ArgM-TMP*(time), for predicate (Verb) *V*. This study assume only complete predicates linked with a sentence so that significant terms and definite predicate (verb) of the sentence are preserved. Predicates are assumed to be complete (appropriate) if they possess at minimum two semantic arguments. If a sentence consists of one predicate, it is expressed by simple PAS, whereas a sentence having more than one predicate is denoted by a composite PAS.

Example 1: Consider the following two sentences represented by simple predicate argument structures.

S_1: "Eventually, a huge cyclone hit the entrance of my house".

S_2: "Finally, a massive hurricane attack my home".

After applying semantic role labeling to sentences S_1 and S^2, the corresponding simple predicate argument structures P_1 and P_2 are extracted are as follows:

P_1: [AM-TMP: Eventually] [A0: a huge cyclone] [V: hit] [A1: the entrance of my house]

P_2: [AM-DIS: Finally] [A0: a massive hurricane] [V: attack] [A1: my home]

When the PASs are acheived, they are decomposed into important words, then stop words are removed. The rest of tokens in PASs are transformed to their root form by employing porter stemming algorithm[27]. Subsequently, POS tagger[24] is used to tag each token of semantic arguments (connected with the predicates), with grammatical roles or Part of Speech (POS) tags. Different POS tags includes *NN* (noun), *V* (verb), *JJ* (adjective) and *RB* (adverb) etc. In this study, PASs are compared with each other based on nouns, verbs, location and time arguments. Thus, we retreive only words from PASs, which are tagged as noun, verb, location, and time. Refering to the PASs in Example 1, they are further processed as shown as follows:

P_1: [*AM-TMP*: *Eventually* (*RB*)] [*A*0: *cyclone* (*NN*)] [*VBD: hit*] [*A*1: *entrance* (*NN*), *house* (*NN*)]

P_2: [*A*0: *hurricane NN*] [*V*: *attack*] [*A*1: *home* (*NN*)]

3.2 Semantic Similarity Matrix

This step builds a similarity matrix from the similarity scores computed for each pair of PAS. Pair wise similarity of the PASs is determined by comparing corresponding nouns, verbs, location and time arguments. Different semantic similarity measures[28] were analyzed and Jiang's measure was found to have closest association with human judgment.

Thus, this work utilizes Jiang measure[25] for determining semantic similarity amongst pairs of predicate argument structures. This measure is one of the information content based measures and considers that each concept existing in the WordNet[29] possess certain amount of information. This measure computes the similarity of given concepts based on the shared information possessed by the concepts. For any given two concepts, Jiang measure[25] uses the following equation to determine their semantic similarity.

$$Jiang_{dist}(C1,C2) = IC(C1) + IC(C2) - 2 \times IC(lso(C1,C2)) \tag{1}$$

Jiang measure utilizes WordNet to calculate the least common subsumer (**lso**) of the given two concepts. lso is the nearest common parent of the given two concepts. The Information Content (IC) of any concept is estimated by computing the likelihood of a concept to occur in a huge text corpus and is given as follows:

$$IC(C)=-logP(C) \tag{2}$$

Whereas is the probability of concept '*C*' to occur and is calculated as follows:

$$P(C) = \frac{Freq(C)}{N} \tag{3}$$

Whereas *Freq*(*C*) is the occurrences of concept '*C*' in the WordNet taxonomy. *N* is the total number of nouns.

Consider are the two sentences, then the semantic similarity of the extracted PASs and is expressed by and is calculated using (8); where is the similarity of predicates (or verbs), computed using (5), is the summation of similarities among corresponding semantic arguments of predicates computedusing Equation (4). Both Equation (4) and Equation (5) use Jiang's measure for determining similarity of nouns in the semantic arguments of the PASs and the verbs of PASs respectively. Similarity score of temporal (time) arguments is computed using Equation (6) and similarity score of location arguments is calculated using Equation (7). Since Jiang measure relies on WordNet, which may not contain the time and location arguments, so the similarity of time and location arguments of the Verb (Predicate) is determined by using edit distance algorithm as given in Equation (6) and Equation (7). The similarity between any two PASs is calculated using Equations (4-8).

$$sim_{arg}(p_i,p_j)=sim(A0_i,A0_j)+sim(A_1 i,A_1 j)+sim(A_2 i,A_2 j) \quad (4)$$

$$sim_{verb}(p_i,p_j)=sim(Verb_i,Verb_j) \quad (5)$$

$$sim_{tmp}(p_i,p_j)=sim(Tmp_i,Tmp_j) \quad (6)$$

$$sim_{loc}(p_i,p_j)=sim(Loc_i,Loc_j) \quad (7)$$

Equations (4), (5), (6), (7) are combined to give Equation (8) as follows:

$$sim_{sem}(p_i,p_j)= sim_{verb}(p_i,p_j)+[sim_{arg}(pi,p_j)+sim_{tmp}(pi,p_j)+sim_{loc}(p_i,p_j)] \quad (8)$$

As the similarity for each pair of PAS is attained, then semantic similarity matrix is built using similarity scores of the PAS and. is described as follows:

$$M_{i,j} = \begin{cases} M_{sim}(p_i, p_j) & if\ i \neq j \\ 0 & else \end{cases} \quad (9)$$

Where represents semantic similarity score of PASs and in the matrix

3.3 Semantic Clustering of PASs

A renowned method in the hierarchical clustering solutions is the Agglomerative Hierarchical Clustering (HAC), which is relatively ancient but proved very helpful and effective in the variety of applications[30]. There exist five familiar linkage methods for (HAC) based on distance measure[31]. These methods include ward's and centroid method and three linkages methods such as single, complete and average linkage method. Various measures (Entropy and F-Score and Kendall W test) were employed in[32–34] and found average linkage as appropriate method for clustering of documents.

Hence, this work utilizes HAC algorithm using average linkage method. The procedure for clustering similar PASs is given as follows.

Pseudo code for Aggloumerative Clustering Algorithm

Input: Semantic Similarity Matrix

Output: Clusters of similar predieate argument structures

The algorithm assumes ij^{th} entry of the matrix as the similarity between i^{th} and j^{th} clusters.

a. Merge the two clusters that are most similar.

b. Update the similarity matrix to reflect the pair wise similarity between the newest cluster and the original cluster based on average linkage method.

c. Repeat step 1 and 2 until the compression rate of summary is reached.

This study assumes 20% as compression rate of genreated summary.

3.4 Selection of PASs using Optimized Features

This step chooses high scored PASs from each cluster using selected features. The optimal weights of these selected features are determined using Particle Swarm Optimization (PSO)[35]. The following features have been selected in this study as they are widely used in text summarization.

3.4.1 Text Features

PAS to PAS Similarity- The semantic similarity between each PAS *P* and the rest of PASs in the document set is determined using Equation (8). This feature is calculated as summation of PAS similarities with the rest of PASs and the maximum PAS similarity in the document set[36].

$$P\_F_1 = \frac{\sum sim(p_i, p_j)}{Max\left(\sum sim(p_i, p_j)\right)} \quad (10)$$

Position of PAS- Location of PAS[37] signifies the salience of the PAS in the document and is to equal to location of sentence. This feature is calculated as follows:

$$P\_F_2 = \frac{document\ length - PAS\ Position + 1}{document\ length} \quad (11)$$

Proper Nouns- The PAS containing more proper nouns are deemed to be salient for summary generation. This feature is determined as follows[37].

$$P\_F_3 = \frac{No.\ of\ \ proper\ nouns\ contained\ in\ PAS}{PAS\ length}$$

(12)

Number of Nouns and Verbs- Composite PAS containing more than one predicates are believed as salient for summary. This feature[36] is determined as follows:

$$P\_F_4 = \frac{No.\ of\ nouns\ and\ verbs\ contained\ in\ the\ PAS}{PAS\ length}$$

(13)

Term Weight- TF-IDF method is commonly used to compute the score of important term [38]. This method is applied to the PAS collection and determine the weights of terms representing nouns and verbs in the PAS. The term weight is determined as follows:

$$W_i = Tf_i \times Idf_i = Tf_i \times \log\frac{N}{n_i} \quad (14)$$

Tf_i is the frequency of the term i in the document, N represents the maximum documents, and n_i indicates the documents containing the term i. This feature is determined as summation of all term weights contained in the PAS and the maximum summation of the term weights in the PAS contained in the document collection[37].

$$P\_F_5 = \frac{\sum_{i=1}^{k} W_i(P)}{Max\left(\sum_{i=1}^{k} W_i(P)\right)} \quad (15)$$

3.4.2 PSO for Optimal Feature Weighs

Text features are cornerstones in the process of producing text summary. Summary quality is susceptible to text features i.e., different features have varied importance in the process of summary generation. Consequently, feature weighting is believed to be vital for the summary to be produced. This study will utilize PSO[35] to obtain optimal weights for features and then optimized features will be utilized in ranking PASs for summary generation. The inspiration to employ PSO for text summarization problem is that it has been found effective in other relevant problems such as data clustering and text classification[39–41]. This specific PSO based experiment is carried out using DUC 2002 data set[42]. PSO is trained and tested on 59 multi-documents (taken from DUC 2002) by employing 10-fold cross validation.

At first, SENNA SRL is used to obtain PAS from the sentence collection in the document set. The scores of text features discussed in Section 3.4.1, are obtained for each PAS contained in the document set, and thus each PAS P is denoted by a vector, which represents the features scores, $P = \{P\_F_1, P\_F_2, \ldots\ldots P\_F_5\}$. Next, the optimal weights for features are found in order to differentiate among salient and less-salient features. Thus, WP refers to features adjusted by its corresponding weights, $WP = \{W_1P\_F_1, W_2P\_F_2, \ldots\ldots W_5P\_F_5\}$. This study will employ binary PSO to determine the optimized weight of each feature. In binary PSO, the position of particle is represented by a binary string. The bit value 1 indicates that the feature is chosen, while the bit value 0 means that the feature is not chosen. The first bit represents the first feature, and the second bit represents the second feature and so on. Figure 2 depicts the structure of particle position. Velocity of particle is expressed in the same manner, where the each bit value is obtained from sigmoid function.

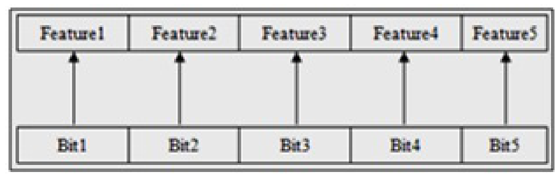

Figure 2. Structure of particle position.

In each iteration of PSO, each particle chooses a particular number of features, and a summary of current multi-document is produced based on selected features. The multi-document summary is then given as input to fitness function. We describe fitness function $F(x)$ as the average recall of multi-document summaries obtained with ROUGE-1[43] and is given in Equation (16).

$$F(x) = \frac{\sum_{S \in \{Reference Summaries\}} \sum_{gramn \in S} Count_{match}(gram_n)}{\sum_{S \in \{Reference Summaries\}} \sum_{gramn \in S} Count(gram_n)}$$

(16)

Where n is the length of the n-gram, the number of *n-grams* that simultaneously occur in system summary and set of human summaries is denoted by , and is the total number of n-grams in the human summaries.

In this study, the number of particles are limited to five as more number of particles can lead to increased computational time[44]. Thus, at the end of each iteration, five evaluation values will appear for five particles. At the first iteration, the pbest for the corresponding particles is determined based on the value obtained from the evaluation of each summary, and the gbest of particle is chosen as the best evaluation value amongst those five summary evaluation values. At the second iteration or above, the new summary evaluation values are compared with the previous pbests to determine the new pbest and gbest of particles. If any new summary evaluation value is superior than the current pbest, then that evaluation value is chosen as pbest. If there is any alteration in the pbest of any particle, then the new pbest will be chosen as gbest if it is better than current gbest. The position of the particle having the gbest value is chosen as vector of best chosen features for the current multi-document at the end of each run. The feature weights $W=\{W_1, W_2........, W_5\}$ for the current multi-document is computed as average of vectors produced in each run. We used 500 maximum generations in order to permit the PSO algorithm to attain convergence. The procedure of PSO is shown in Figure 3.

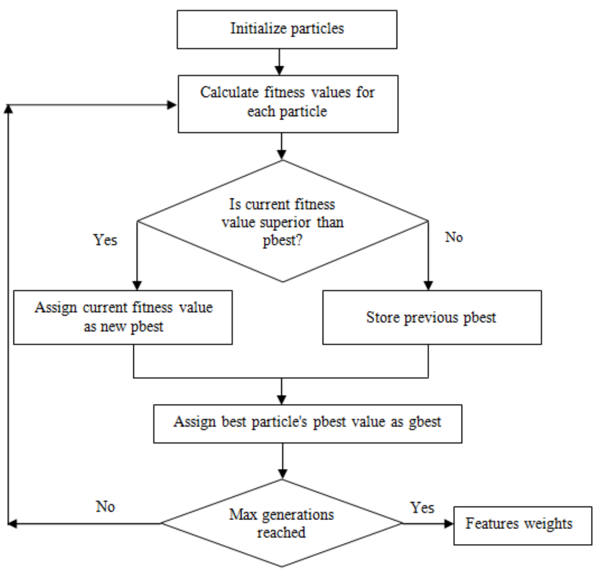

Figure 3. Procedure of Particle Swarm Optimization.

The final vector of optimal features weights is determined from the vectors representing features weights for all multi-documents in the data set. The optimal set of feature weights achieved with PSO are employed to adjust the features (extracted for predicate argument structures) by their corresponding weights. The score of predicate argument structure adjusted by feature weights, is computed as follows:

$$Adjuste\_Score(\Pr_i) = \sum_{k=1}^{5} W_k \times \Pr\_F_k(\Pr_i) \qquad (17)$$

Where is the predicate argument structure score based on optimized features, represents feature k score for predicate argument structure , and is the optimized weight obtained for feature k.

Thus, the optimal features are used to score the PAS in each cluster. After the scores of PASs are obtained, the PASs in all clusters are represented by their corresponding scores, and arranged in descending order. The predicate argument structure having highest rank is picked from each cluster. The selected top ranked PASs are fed to the next phase.

3.5 Abstractive Summary Generation

In this step, the high scored representative PASs are taken from previous phase. It makes use of SimpleNLG[26] embedded with heuristic rules to produce sentences from PASs.

SimpleNLG engine offers simple interfaces to construct syntactical structures and exploits simple grammar rules to transform these structures into sentences using. Furthermore, the engine is robust i.e., the it will not collapse, when incomplete or ill-formed syntactical structures are given as input.

The first heuristic rule states that *"if the subjects in the predicate argument structures (PASs) refer to the same entity, then merge the predicate argument structures by removing the subject in all PASs except the first one, separating them by a comma (if there exist more than two PASs) and then combine them using connective 'and'* ".

The second rule states that *"If PAS P_i subsumes a PAS P_j, then the subsumed PAS P_j is discarded in order to avoid redundancy"*.

As discussed in Section 3.1, we consider specific arguments i.e., the *core arguments: A0 (subject), A1 (object), A2 (indirect object), and adjunctive arguments: ArgM-LOC (location), ArgM-TMP(time), for predicate (Verb) V*. So, the sentences produced from the given PASs will be the compact from of the source sentences in many cases. The heuristic rules embedded in SimpleNLG combines the PASs that represent the matching subject. The given example shows how summary is produced from the source sentences.

We consider that the top ranked PASs chosen from previous step are P_1, P_2 and P_3. Keeping in view the rule and example mentioned above, the subject *A0* is found as replicated in all PASs and is removed from all PASs excluding the first one. The first heuristic rule is performed on all PASs and constitute the summary sentence, which is the compact form of source sentences.
For instance, the following source input setences:
S_1: *"Hurricane Gilbert claimed to be the most intense storm*

on record in terms of barometric pressure".
S_2: *"Hurricane Gilbert slammed into Kingston on Monday with torrential rains and 115 mph winds".*
S_3: *"Hurricane Gilbert ripped roofs off homes and buildings".*
After applying SENNA SRL, the corresponding three predicate argument structures P1, P2 and P3 are obtained as follows:
P1: [A0: Hurricane Gilbert] [V: Claimed] [A1: to be the most intense storm on record]
P2: [A0: Hurricane Gilbert] [V: slammed] [A1: into Kingston] [AM-TMP: on Monday]
P3: [A0: Hurricane Gilbert] [V: ripped] [A1: roofs off homes and buildings]

We consider that the top ranked PASs chosen from previous step are P1, P2 and P3. Keeping in view the rule and example mentioned above, the subject A0 is found as replicated in all PASs and is removed from all PASs excluding the first one. The first heuristic rule is performed on all PASs and constitute the summary sentence, which is the compact form of source is performed on all PASs and constitute the summary sentence, which is the compact form of source sentences.

Summary Sentence: "Hurricane Gilbert claimed to be the most intense storm on record, slammed into Kingston on Monday with torrential rains and ripped roofs off homes and buildings".

4. Evaluation Results

The optimized semantic technique for MDAS is assessed using DUC 2002 data sets[42], which is a benchnmark corpus widely employed in text summarization community. The corpus includes 59 document sets along with the human made abstractive summaries, created by the National Institute of Standards and Technology (NIST). The data set chosen for this study refers to task3 defined for DUC 2002.

The two evaluation metrics, ROUGE[43] and Pyramid[45] have been extensively employed for evaluation text summaries. ROUGE computes the score based on n-gram matches found in system summary and human summaries. It can not recognize the meanings of words. The pyramid metric is employed for the abstractive summary evaluation. The strength of Pyramid metric is that it can recognize diverse sentences in the summaries that are semnatically equivalent[45].

We employ Pyramid evaluation results to compare our optimized semantic technique (SRL-PSO) with the recent abstractive approach for MDS (AS)[14], the best system, average of automatic systems, and average of human (model) summaries, in the perspective of DUC 2002 shared tasks for MDAS.

Mean Coverage Score[45] for peer summary is determined as given below:

$$MeanCoverageScore(MCS) = \frac{Total\,weight\,of\,Peer\,SCUs}{Average\,SCUs\,in\,Human\,Summary} \quad (18)$$

SCUs represent summary content units and their weights refers to the number of human summaries containing SCUs.

Precision of peer summary[45] is calculated as given below.

$$\Pr ecision(P) = \frac{Model\,SCUs\,contained\,in\,Peer\,Summary}{Average\,SCUs\,in\,Peer\,Summary} \quad (19)$$

The F-measure for the given peer summary is determined from Equation (19) and Equation (20) as follows:

$$F - Measure = \frac{2 \times MCS \times P}{MCS + P} \quad (20)$$

For each data set amongst the 59 news articles/data, our proposed technique produces a 100 words summary, the task performed by other systems that participated in MDAS tasks. To compare the effectiveness of the proposed technique (SRL-PSO) in the perspective of DUC 2002 shared tasks for MDAS, we establish four summarization models (Best, Avg, AS-SRL, AS), in addition to average of human summaries (referred to as Models). Table 1 depicts the results of different summarization models over three evaluation metrics (mean coverage score, precision and F-measure) tested on DUC 2002 dataset.

Table 1. Comparative evaluation of differnt summarization models based on mean coverage score, precision and F-Measure

System	Mean Coverage Score	AVG-Precision	AVG-F-Measure
DUC-2002 Models	0.6910	0.8528	0.7634
AS-PSO-SRL	0.4732	0.6943	0.5628
AS-SRL	0.4397	0.60	0.5078
AS [14]	0.4378	0.643	0.5209
DUC-2002 Best (System 19)	0.2783	0.7452	0.4053
DUC-2002 Avg	0.1775	0.6700	0.2806

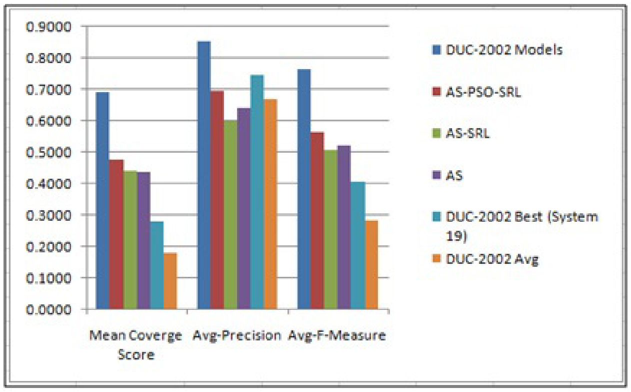

Figure 4. Comparison of summarization results based on mean coverage score, average precision and average F-measure.

The optimal weights for different featurse obtained using PSO are depicted in Figure 5. The optimal weights obtained are 0.40314, 0.67179, 0.33148, 0.56478, 0.46556 correspond to the features including PAS to PAS semantic similarity, position, proper nouns, nouns and verbs, and TF_IDF respectively.

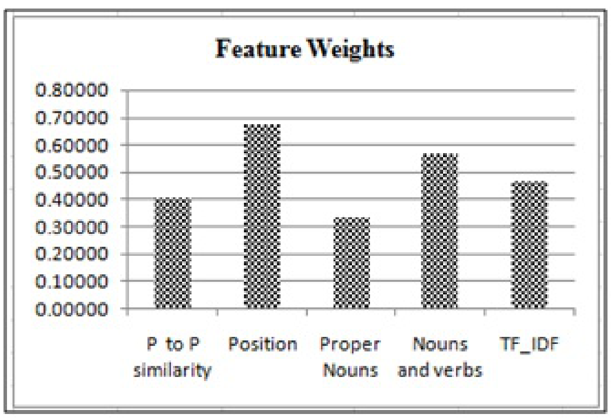

Figure 5. Optimal feature weights obtained using PSO.

4.1 Discussion

In this section, we give details of the results given in preceding section. Table 1 compares the evaluation results of proposed technique and the benchmark summarization models based on evaluation measures: Mean coverage score, precision and F-measure, in the perspective of DUC 2002 shared tasks for MDAS. It could be seen from the results in Table 1 that on mean coverage score and average F-measure, our semantic approach (AS-PSO-SRL) yields better summarization results than other comparison summarization models; and appeared next to the Models, which represents average of human summaries. On the other hand, on precision measure, our approach performed slightly inferior than Best system but yet perfromed superior than other approaches; and stood third to the Models. Summarization results obtained with our technique and the rest of comparison models are validated by carrying out a Paired-Samples T-test, which yields a lesser significance value i.e., $p < 0.05$. The achieved results verify that summary generated by proposed technique (AS-PSO-SRL) is more closer to human's summary while testing against the comparison models (AS-SRL, AS, Best and Avg).

In order to examine the influence of PSO on summarization, the PSO module is dropped from the proposed summarization approach, and we called it AS-SRL, which presume that all features hold same importance. It could be observed from Table 1, that the performance of the technique without integration of PSO (AS-SRL) degrades on mean coverage score, precision and F-measure. A statistical significance tests (*T*-Tests) is also carried out to in order to reveal the improvement of the proposed method (AS-PSO-SRL) over other summarization model (AS-SRL). The low significance value achieved for the *T*-test (typically less than 0.05) reveals that the summarzation results obtained with proposed technique (AS-PSO-SRL) and comparison model (AS-SRL) are significantly different. These results suggest that incorporation of PSO into the proposed abstractive summarization approach enhanced the summarization results.

5. Conclusion

Abstractive summarization is an emerging area for researchers, our proposed technique illustrates the viability of this novel track for summarization research. Experimental results reveal that the optimized semantic technique yields better results than benchmark comparison models and appeared next to the humans.

Furthermore, PSO is utilized in the proposed technique to find out the relevance of each feature by giving them suitable weights, and later the optimal feature weights are utilized to choose the top scored PASs for summary to be generated. The results reveal that integration of PSO in the proposed technique improved summarization results when evaluated the technique without PSO. In future, we plan to integrate fuzzy logic in the proposed technique to flexibly tolerate the uncertain, imprecise and ambiguous feature weights.

6. Acknowledgement

This work is supported by the Higher Education Commission (HEC), Islamia College Peshawar (Chartered University), Pakistan and Soft Computing Research Group (SCRG) of Universiti Teknologi Malaysia (UTM). This Work is also supported in part by grant from R.J130000.7828.4F719.

7. References

1. Khan A, Salim N, Kumar YJ. A framework for multi-document abstractive summarization based on semantic role labelling. Applied Soft Computing. 2015; 30(2015):737–47.
2. Kumar YJ, Salim N, Abuobieda A, Tawfik A. Multi document summarization based on cross-document relation using voting technique. 2013 IEEE International Conference on Computing, Electrical and Electronics Engineering (ICCEEE); 2013. p. 609–14.
3. Das D, Martins AF. A survey on automatic text summarization. Literature Survey for the Language and Statistics II Course at CMU. 2007. p. 4192–5.
4. Ye S, Chua T-S, Kan M-Y, Qiu L. Document concept lattice for text understanding and summarization. Information Processing and Management. 2007; 43(6):1643–62.
5. Kupiec J, Pedersen J, Chen F. A trainable document summarizer. Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval; Washington, USA. 1995 Jul 9-13. p. 68–73.
6. Gayathri P, Jaisankar N. An efficient medical document summarization using sentence feature extraction and ranking. Indian Journal of Science and Technology. 2015; 8(33).
7. Knight K, Marcu D. Statistics-based summarization-step one: Sentence compression. Proceedings of the 17th National Conference on Artificial Intelligence; Austin, Texas: AAAI Press. 2000 Jul 30-August 3. p. 703–10.
8. Prakash S, Chakravarthy T, Brindha G. Preference based quantified summarization of on-line reviews. Indian Journal of Science and Technology. 2014; 7(11):1788–97.
9. Larsen B. A trainable summarizer with knowledge acquired from robust NLP techniques. Advances in Automatic Text Summarization; 1999. p. 71–9.
10. Fattah MA. A hybrid machine learning model for multi-document summarization. Applied Intelligence. 2014; 40(4):592–600.
11. Barzilay R, McKeown KR. Sentence fusion for multidocument news summarization. Computational Linguistics. 2005; 31(3):297–328.
12. Titov I, Klementiev A. A Bayesian approach to unsupervised semantic role induction. ACL Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics; Avignon, France. 2012 Apr 23-27. p. 12–22.
13. Barzilay R, McKeown KR, Elhadad M. Information fusion in the context of multi-document summarization. ACL Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics; College Park, Maryland. 1999 Jun 20-26. p. 550–7.
14. Genest P-E, Lapalme G. Framework for abstractive summarization using text-to-text generation. ACL Proceedings of the Workshop on Monolingual Text-To-Text Generation; Oregon, USA. 2011 Jun 24. p. 64–73.
15. Tanaka H, Kinoshita A, Kobayakawa T, Kumano T, Kato N. Syntax-driven sentence revision for broadcast news summarization. ACL Proceedings of the 2009 Workshop on Language Generation and Summarisation; Suntec, Singapore. 2009 Aug 6. p. 39–47.
16. Harabagiu SM, Lacatusu F. Generating single and multi-document summaries with gistexter. Document Understanding Conferences; Pennsylvania, USA: NIST. 2002 Jul 11-12. p. 40–5.
17. Lee C-S, Jian Z-W, Huang L-K. A fuzzy ontology and its application to news summarization. IEEE Transactions on Systems, Man and Cybernetics. Part B: Cybernetics. 2005; 35(5):859–80.
18. Genest P-E, Lapalme G. Fully abstractive approach to guided summarization. ACL Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers; Jeju Island, Korea. 2012 Jul 8-14. p. 354–8.
19. Greenbacker CF. Towards a framework for abstractive summarization of multimodal documents. Greenbacker CF, editor. ACL Proceedings of the ACL 2011 Student Session. USA. 2011. p. 75–80.
20. Moawad IF, Aref M. Semantic graph reduction approach for abstractive text summarization. 2012 IEEE 7th International Conference on Computer Engineering and Systems (ICCES); 2012. p. 132–8.
21. Cheung JCK, Penn G. Towards robust abstractive multi-document summarization: A caseframe analysis of centrality and domain. ACL; 2013. p. 1233–42.
22. Collobert R, Weston J, Bottou L, Karlen M, Kavukcuoglu K, Kuksa P. Natural language processing (almost) from scratch. The Journal of Machine Learning Research. 2011; 12:2493–537.
23. Jiang JJ, Conrath DW. Semantic similarity based on corpus statistics and lexical taxonomy. arXiv preprint cmp-lg/9709008; 1997.
24. Bing L, Li P, Liao Y, Lam W, Guo W, Passonneau RJ. Abstractive multi-document summarization via phrase selection and merging. ACL Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing; Beijing, China. 2015 Jul 26-32. p. 1587–97.
25. Boudin F, Mougard H, Favre B. Concept-based summarization using integer linear programming: from concept pruning to multiple optimal solutions. ACL Conference on Empirical Methods in Natural Language Processing (EMNLP); Lisbon, Portugal. 2015 Sep 17-21. p. 1914–8.

26. Gatt A, Reiter E. SimpleNLG: A realisation engine for practical applications. ACL Proceedings of the 12th European Workshop on Natural Language Generation; Athens, Greece. 2009 Mar 30-31. p. 90–3.
27. Porter MF. Snowball: A language for stemming algorithms; 2001.
28. Li Y, Bandar ZA, McLean D. An approach for measuring semantic similarity between words using multiple information sources. IEEE Transactions on Knowledge and Data Engineering. 2003; 15(4):871–82.
29. Miller GA. WordNet: A lexical database for English. Communications of the ACM. 1995; 38(11):39–41.
30. Murtagh F, Contreras P. Methods of Hierarchical Clustering. arXiv preprint arXiv:1105.0121; 2011.
31. Takumi S, Miyamoto S. Top-down vs bottom-up methods of linkage for asymmetric agglomerative hierarchical clustering. IEEE International Conference on Granular Computing (GrC); 2012. p. 459–64.
32. Steinbach M, Karypis G, Kumar V. A comparison of document clustering techniques. KDD Workshop on Text Mining; Boston, USA. 2000 Aug 20-23. p. 525–6.
33. Zhao Y, Karypis G, Fayyad U. Hierarchical clustering algorithms for document datasets. Data Mining and Knowledge Discovery. 2005; 10(2):141–68.
34. El-Hamdouchi A, Willett P. Comparison of hierarchic agglomerative clustering methods for document retrieval. The Computer Journal. 1989; 32(3):220–7.
35. Eberhart RC, Shi Y. Particle swarm optimization: Developments, applications and resources. IEEE Proceedings of the 2001 Congress on Evolutionary Computation; 2001. p. 81–6.
36. Fattah MA, Ren F. GA, MR, FFNN, PNN and GMM based models for automatic text summarization. Computer Speech and Language. 2009; 23(1):126–44.
37. Suanmali L, Salim N, Binwahlan MS. Fuzzy logic based method for improving text summarization. arXiv preprint arXiv:0906.4690; 2009.
38. Salton G. Automatic text processing: The Transformation, Analysis, and Retrieval of information by computer. Addison-Wesley; 1989.
39. Van der Merwe D, Engelbrecht AP. Data clustering using particle swarm optimization. IEEE 2003 Congress on Evolutionary Computation, CEC'03; 2003. p. 215–20.
40. Cui X, Potok TE, Palathingal P. Document clustering using particle swarm optimization. IEEE Proceedings on Swarm Intelligence Symposium, SIS; 2005. p. 185–91.
41. Ziegler C-N, Skubacz M. Content extraction from news pages using particle swarm optimization on linguistic and structural features. Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence: IEEE Computer Society; 2007. p. 242–9.
42. DUC. Document Understanding Conference 2002; 2002.
43. Lin C-Y. Rouge: A package for automatic evaluation of summaries. Proceedings of the ACL-04 Workshop on Text Summarization Branches Out; Barcelona, Spain. 2004 Jul 25-26. p. 74–81.
44. Kiran R, Jetti SR, Venayagamoorthy GK. Online training of a generalized neuron with particle swarm optimization. IEEE International Joint Conference on Neural Networks, IJCNN'06; 2006. p. 5088–95.
45. Nenkova A, Passonneau R. Evaluating content selection in summarization: The pyramid method. NAACL-HLT; 2004.