Research Reviews

1. *Outline*

For every research paper reviewed, write 4 paragraphs outlining: the aim of the paper, the methodology used by the authors in detail, experiments carried out, evaluation of the results, conclusions.

1. Extractive Text Summarization Using Deep Learning – Serry

This paper proposes an extractive text summarization approach utilizing a deep learning model. The approach involves a mixture of engineering and extracting lower level features based on methods that are available in wider literature, latent features are then extracted using a restricted bolts man machine to enhance the quality of the generated summaries without any loss of information. Results are compared to existing approaches that utilize RBM’s

The proposed approach involves multiple phases, the first phase *‘Preprocessing’* involves segmenting documents in the corpus into paragraphs to keep track which sentence belongs to which paragraph and the position of a sentence in its respective paragraph. Paragraphs are further divided into sentences followed by breaking down each sentence into words that are lemmatized, stop words are also removed from sentences. Lastly, each word is tagged with respect to its grammatical identifier i.e. noun, verb…etc.

the second phase *‘Feature Extraction’* involves engineering features that would be utilized to enhance overall quality of summaries generated. The features selected by the author are as follows: **number of thematic words** the 10 most frequently recurring words in a text, **sentence position**, **sentence length, sentence position relative to paragraph**, this comes from an observation by the author that a new discussion begins at the start of each paragraph. The next feature proposed is **number of proper nouns per sentence**, **number of numerals**, **number of named entities**, **TF-ISF** which is like the TF-IDF weighting scheme but tailored to one document only. Lastly **sentence-centroid similarity**, the sentences with the highest TF-ISF score is considered a centroid and all other sentences are compared to it.

The last phase is *‘Feature Enhancement’* by the end of the second phase a sentence-feature matrix is generated. Each sentence is represented as feature vector of length 9. Each sentence vector is then given as input to an RBM with 1 hidden layer and 1 visible layer where each sentence feature is multiplied by the learned weights in the RBM resulting in a more enhanced sentence-feature matrix. The RBM is trained for each new document. The last step is the generation of a summary, the sentence feature vectors are summed to give a final score for each sentence. Sentences with the highest score are used for the summary.

For experimentation, the author obtained factual reports from several domains such as health, technology. The approach mentioned above was applied on each report and the generated summaries were compared to those produced by human experts. The evaluation of results is carried out based on the following metrics: F-measure, recall, precision. The results were then compared to an ‘existing approach’ that uses RBMs to indicate a fair improvement in performance.

In Conclusion, the paper provides a reasonable approach to summarization. However, the author remarks that existing approaches utilize 2 or more layers. He further remarks that 1 hidden layer would suffice given the size of the training data available. Many such ‘given conclusions’ are to be observed through out the paper without clear evidence. Furthermore, the existing approaches he mentions are no where to be found in the paper making it difficult to evaluate the results.

1. Extractive Text Summarization Using Supervised and Semi-supervised Learning – Serry