Extractive Summarization using Deep Learning

This paper proposes a text summarization approach for factual reports using a deep learning model. This approach consists of three phases: feature extraction, feature enhancement, and summary generation, which work together to assimilate core information and generate a coherent, understandable summary. We are exploring various features to improve the set of sentences selected for the summary, and are using a Restricted Boltzmann Machine to enhance and abstract those features to improve resultant accuracy without losing any important information. The sentences are scored based on those enhanced features and an ex-tractive summary is constructed. Experimentation carried out on several articles demonstrates the effectiveness of the proposed approach.

Pre-processing:

1. Document Segmentation:
2. Paragraph Segmentation:
3. Word Normalization:
4. Stop Word Filtering:
5. PoS Tagging:

Feature Extraction:

1. Number of thematic words:
2. Sentence position:
3. Sentence length:
4. Number of proper nouns:
5. Number of numerals:
6. Term Frequency-Inverse Sentence Frequency (TF ISF):

Conclusion:

We have developed an algorithm to summarize single-document factual reports. The algorithm runs separately for each input document, instead of learning rules from a corpus, as each document is unique in itself. This is an advantage that our approach provides. We extract 6 features from the given document and use them to score each sentence.

Input :

<https://en.wikipedia.org/wiki/National_Institute_of_Technology,_Patna>

Output:

The National Institute of Technology Patna (NIT Patna), formerly Bihar School of Engineering and Bihar College of Engineering, is a public engineering institution located in Patna in the Indian state of Bihar. It was renamed to NIT Patna, by the Government of India on 28 January 2004. It is an autonomous institute functions directly under Ministry of Human Resource Development, Government of India. NIT Patna origin can be traced to 1886 with the establishment of a survey training school and subsequent renaming it to Bihar college of Engineering in 1900. A graduate level curriculum was introduced in 1924. It was renamed Bihar College of Engineering in 1932. In 2004 the government of India upgraded the college to National Institute of Technology (NIT) status, as the state of Bihar had lost its only Regional Engineering College (REC), located at Jamshedpur, when Jharkhand was carved out of Bihar in 2000.