

## 1.1 Citation - 1 point per paper

Jianpeng Cheng and Mirella Lapata (2016). Neural Summarization by Extracting Sentences and Words. *In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*.

## 1.2 Task - 2 points per paper

Jianpeng and Mirella (2016) proposed a data driven approach based on neural networks and continuous sentence features for *extractive text summarization*. In contrast, the traditional approaches rely heavily on human engineered features.

## 1.3 Data - 2 points per paper

Jianpeng and Mirella (2016) collected the hundreds of thousands of news article with their corresponding highlights from DailyMail for both sentence and word extraction. For word extraction, building the dataset was straight forward (searching for words in the document and highlight). In case of sentence extraction, as the highlights were written by real humans, they are better suited for abstractive text summarization, so Jianpeng and Mirella (2016) reverse approximated the gold standard label of each document sentence given the summary based on their semantic correspondence.

## 1.4 Approach - 5 points per paper

Jianpeng and Mirella used a neural network based *hierarchical* document reader and an attention based hierarchical content extractor. Let's briefly describe the document reader and content extractor (sentence and word extractor) one by one.

The document reader consists of:

- a) single layer CNN with a max-over-time pooling operation to obtain a representation vector at the sentence level. It uses a tanh activation function and perform a max pooling over time to obtain a single feature.
- b) a standard RNN that recursively composes sentences to build representation for documents. It uses an LSTM activation unit.

The content extractor is another RNN that labels sentences sequentially also taking into account for any redundancy. The output is an extractive summary of the document.

## 1.5 Evaluation - 5 points per paper

Jianpeng and Mirella used a task based evaluation in which the human judges were presented with a news article and summaries generated by a list of other systems. The articles were 20 randomly sampled DUC 2002 test documents. The other contenders were the LEAD baseline, the phrase-based ILP model and the human authored summary. The participants were asked to rank the summaries best to worst in order of informativeness and fluency. The Jianpeng and

Mirella model outperformed the LEAD with a significant margin while performing slightly better than the ILP model. The LEAD uses a set of manually selected features while the ILP takes advantage of syntactic information and extract summaries subject to well engineered linguistic constraints, which Jianpeng and Mirella's model lacks.

Also, the ROUGE score for the Jianpeng and Mirella's model outperformed all other models in the ROUGE - 1, 2 (as a means of assessing informativeness) and the LCS (ROUGE - L) as a means of assessing fluency on both DailyMail and DUC 2002 test documents.