Advanced NLP

Reference Projects List

I. Domains:

1. Machine Translation
2. Question Answering
3. Summarization
4. Conversational Systems

II. Project Descriptions

1. Machine Translation of News Articles on ACL 2019 shared task dataset

2. Implement Rank QA : Neural Question Answering with Answer Re-Ranking paper. And compare with other state-of-art papers of your choice. Also compare your results on different datasets like: SQUAD , WIKI

Link to Paper : https://www.aclweb.org/anthology/P19-1611.pdf
Baseline: Implement the paper on SQUAD and WIKI data sets
Further Improvements: Compare with other state-of-art papers Question and Answering papers of your choice(minimum 2)

3. : Scientific Document Summarization shared task https://wing.comp.nus.edu.sg/~cl-scisumm2019/

4. Given a paragraph/document in english wiki, 'translate' it into simple wiki - Should work even for docs out of wiki.

Methods/Variants : Statistical MT Approach, Encoder-Decoder/WordEmbeddings
Dataset and description : http://www.cs.pomona.edu/~dkauchak/simplification/ (Aligned data)

5. Implement a state-of-art paper on Community Question and Answering and Propose your improvements over the baseline model

Datasets: Yahoo answers

Baseline: Implement a state-of-art paper of your choice for baseline

Further Improvements: Suggest improvements over the baseline and implement them

6. Domain Term Extraction

Papers:

1. An Unsupervised Approach to Domain-Specific Term Extraction
2. Term extraction using non-technical corpora as a point of leverage
3. Domain-Specific Term Extraction and Its Application in Text Classification

Dataset: Prepare Dataset by web scraping Wikipedia

Baseline: Collect Data from Wikipedia and implement a model (either from one of the papers or a hybrid model) . The goal is , on a new document , we should be able to identify the Domain Terms

Further Improvements: Improve the Data , Come up with suggestions on your Baseline to improve the results.

7,. Goal : Implement State-of-art papers on Open Domain Question and Answering

Dataset: WikiQA dataset

Baseline: Implement a paper of your choice as the Baseline model.

Further Improvements: Explore more papers in the same area and come up with an improved model of your baseline model

8. Goal : Scientific Document Summarization shared task
https://wing.comp.nus.edu.sg/~cl-scisumm2019/

9. To incorporate the benefits of multiple MT systems into one, so as to improve upon the performances of the individual baseline systems.

10. NLP for Social Media. Hate Speech, Code mix tasks

11. Argument Mining: Detect Arguments and claims in unstructured data

Baseline: Sequence Labelling on Essays Dataset

Baseline++:Relation prediction (for/against) between premises and claims

12. Bias Detection in news articles. Detect sentence level and article level bias in news domain.

13. Semantic Textual Similarity: To address the problem of semantic coincidence between sentence pairs. Commonly knownas paraphrase identification.

14. Natural Language Inference : To understand semantic concepts like textual entailment and contradiction. The task isthat of comparing two sentences and identifying the relationship between them.