Discover & Learn (1-2-3-4)

Text Extraction, Simplification and Synopsis Generation using Reinforcement Learning

Under the supervision of Prof. Niladri Chatterjee

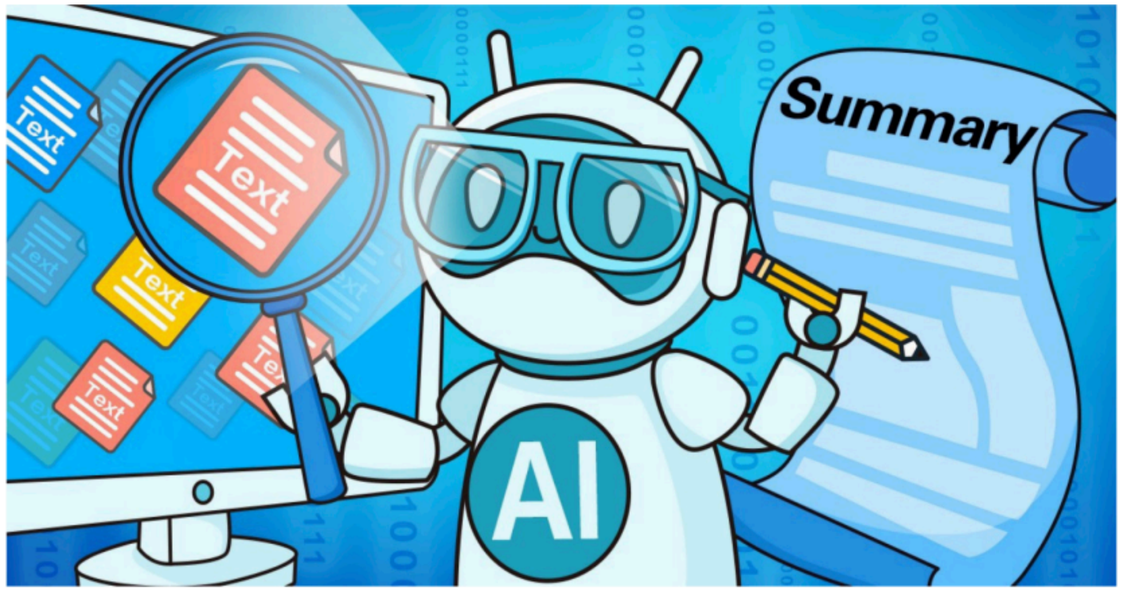
Bhumika Chopra (2018MT10748)
Hetvi Jethwani (2018MT10754) Mallika Prabhakar (2019CS50440) Sakshi Taparia (2017MT10748) Sharut Gupta (2017MT60250)

MI Number- **MI02130G , MI02126N**

Scheme : **D&L - 2019**

Serial Number - **17**

Our aim is to generate a meaningful summary from a given report using Natural Language Processing and Reinforcement Learning techniques.



Motivation and Applications?

1. Analyse huge volumes of corporate reports
2. Promote efficient understanding and usage
3. Simplify and reduce the text for people with cognitive disabilities

Why Reinforcement Learning?

1. Abstractive Summarisation
2. Generation and Compression tasks

distribution over
next states given
current

Environment
(MDP)

Probability
state and action Prescribes action to take for each
state

Reinforcement

Reward Optimal Policy

Learning being in a state

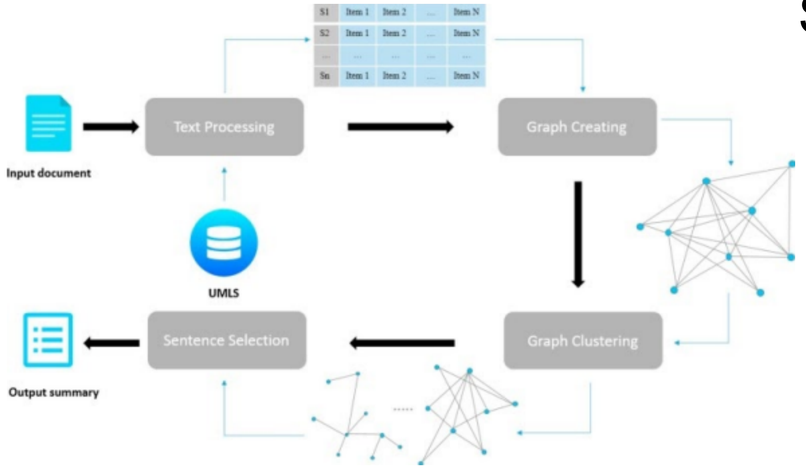
or Trajectories

Describes
desirability of

The algorithm takes suitable action to maximize reward in a particular situation!

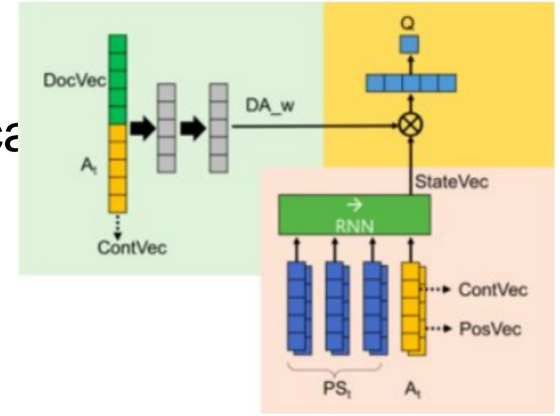
Approaches.

Graph Based



Examples: LexRank

Statistica

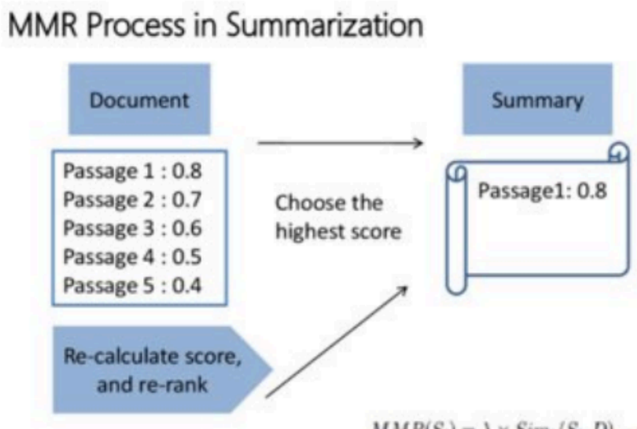


Examples:

RELIS
REFRESH
DRESS
LEAD

NN and RL Based

Graph si

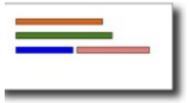


Examples:

Cue word TF - IDF RST
MMR

State of the art Methods!

Baseline



Statistical (MMR) Method

; Dataset

18.5 5.8 11.2

15.9 4.6 10.9

11.2 3.5 7.7

12.1 4.4 9.9

Extractive
Approach

Experimenting with word embeddings

Graph based Method

Text Summarisation

20.2 4.9 13.1

20.3 5.1 13.2

12.9 3.0 10.2

19.4 4.8 12.6

24.4 6.9 14.5

24.39 6.98 14.47 **28.17 8.7 19.72** **29.1 11.1 25.9**

30.4 11.7 26.9

Performance of conventional algorithms is highly dependent on word embeddings used. As expected, the RL based methods improve the results significantly.

Run all current and future experiments on Nokia dataset.

Extend applicability of methods from single sentence to paragraph, document, and collection of documents.

Develop suitable evaluation and performance measuring metrics.

Work Plan

Year 1	Study and implement the basic reinforcement learning models for text simplification. Study the applicability and suitability of word embeddings with respect to the documents provided by Nokia.
Year 2	Retrain the model based on word embedding schemes.
Year 3	Extending the applicability of the text simplification algorithm from a single sentence to a paragraph, a document and a collection of documents.
Year 4	Development of suitable evaluation and performance measuring scales. Searching over future possibilities.

Thank you!

Questions?