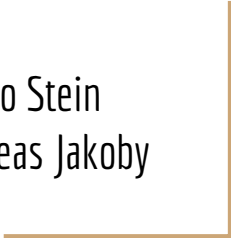# Learning Unified Multi-Document Summarization From Collaborative Journalism

Master's Thesis by Yasar Naci Gündüz

First Referee : Prof.Dr.Benno Stein
Second Referee : Prof.Dr.Andreas Jakoby

INTRODUCTION: New age, new habits

INTRODUCTION: New age, new habits

Introduction : How about journalism?

Several research reported:

- Reading attention span is getting shorter
- Young generation is the least informed…
- …and more interested in social media

Introduction : How about journalism?

Several research reported:

- Reading attention span is getting shorter
- Young generation is the least informed...
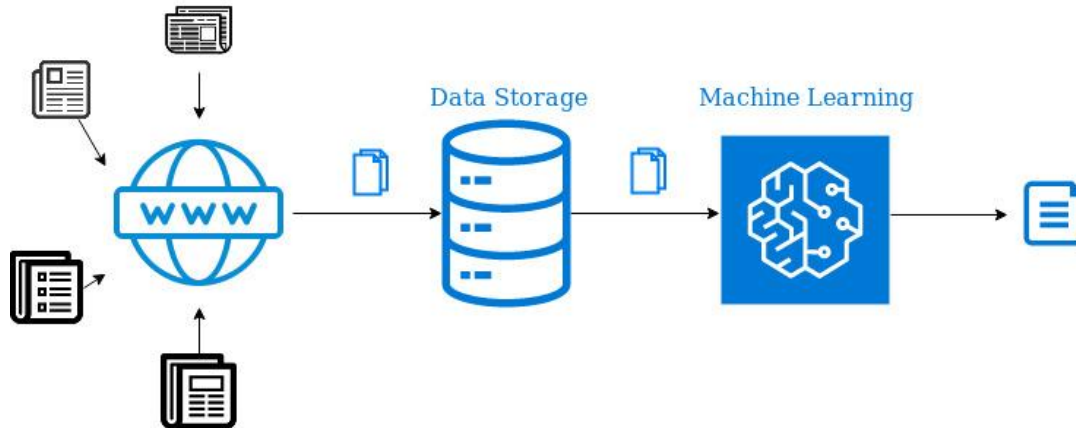- ...and more interested in social media

Information Pollution:

- Reliable sources are more important than ever

Introduction : Our proposal

Make the content:

- Less time consuming
- Yet still adequately informing

Solution: Automatic Summarization

Introduction : Automatic Summarization for Journalism

"Journalism is the activity of gathering, assessing, creating, and presenting news and information."

American Press Institute

Introduction : Automatic Summarization for Journalism

"Journalism is the activity of gathering, assessing, creating, and presenting news and information."

- Whole
- Extensive
- Unbiased

Introduction : Automatic Summarization for Journalism

"Journalism is the activity of gathering, assessing, creating, and presenting news and information."

- Whole
- Extensive
- Unbiased

Solution: Multi-document Summarization

Introduction : Automatic Summarization for Journalism

"Journalism is the activity of gathering, assessing, creating, and presenting news and information."

- Extractive and Abstractive

Introduction : Automatic Summarization for Journalism

"Journalism is the activity of gathering, assessing, creating, and presenting news and information."
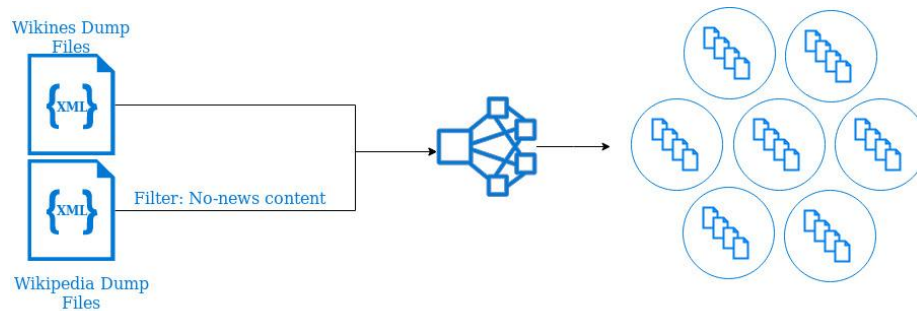
- Extractive and Abstractive
- Neural Abstractive Summarization
  - Methods are generally for Single-Document

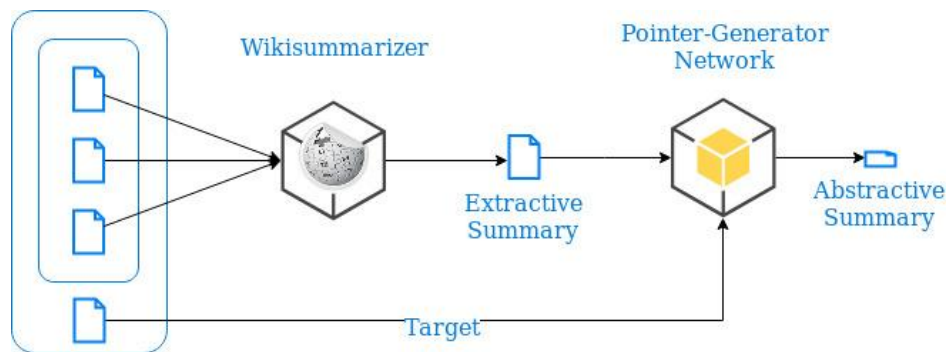# Introduction : Automatic Summarization for Journalism

"Journalism is the activity of gathering, assessing, creating, and presenting news and information."

- Extractive and Abstractive
- Neural Abstractive Summarization
  - Methods are generally for Single-Document
- Unified Model : Extractive + Abstractive
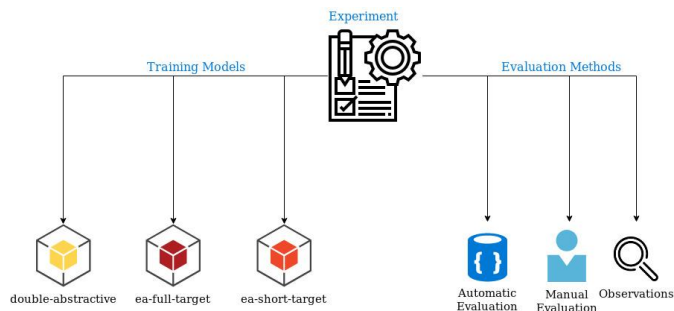 - Content Selection
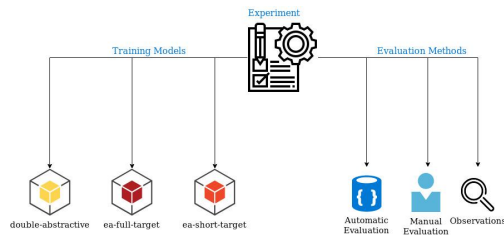 - Multi-Document -> Single Document

Dataset

Unified Summarization Pipeline

Experiments&Evaluation

13

Dataset

Wikines Dump Files

{XML}

Filter: No-news content

{XML}

Wikipedia Dump Files

Wikisummarizer

Pointer-Generator Network

Extractive Summary

Abstractive Summary

Target

Experiment

Training Models

Evaluation Methods

double-abstractive   ea-full-target   ea-short-target

Automatic Evaluation   Manual Evaluation   Observations
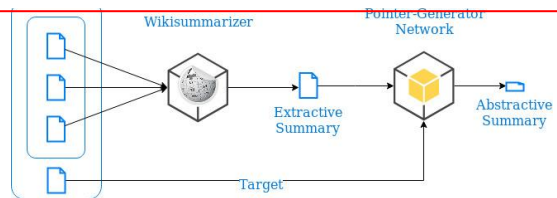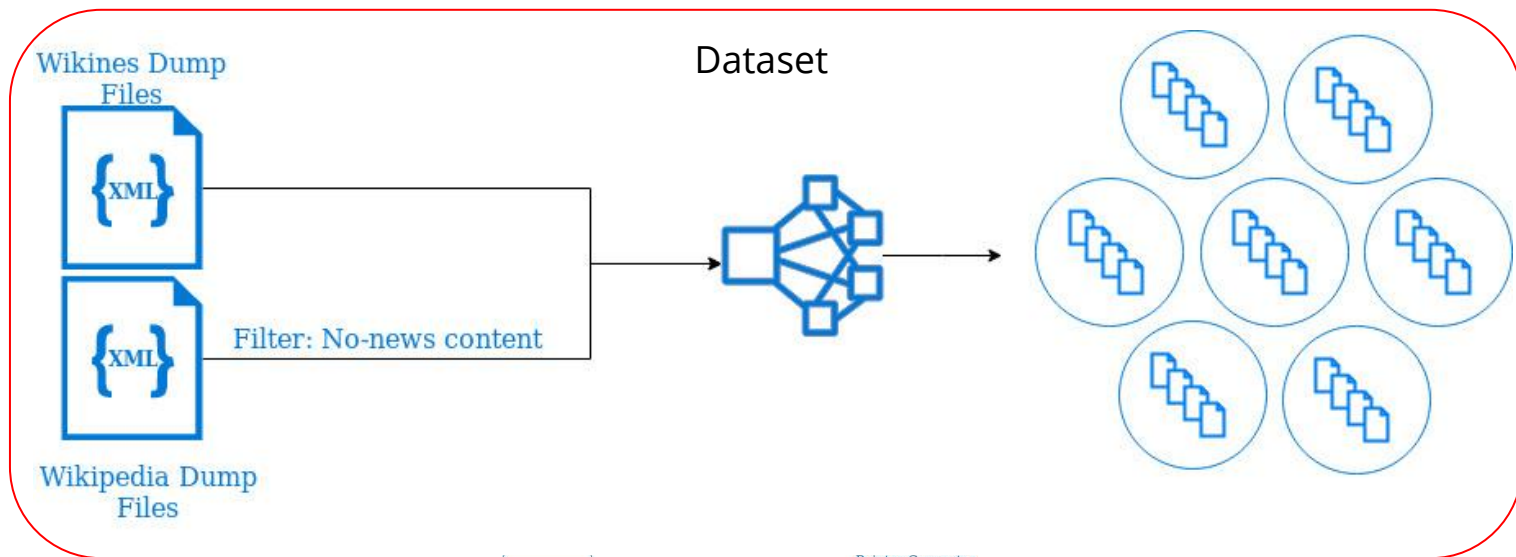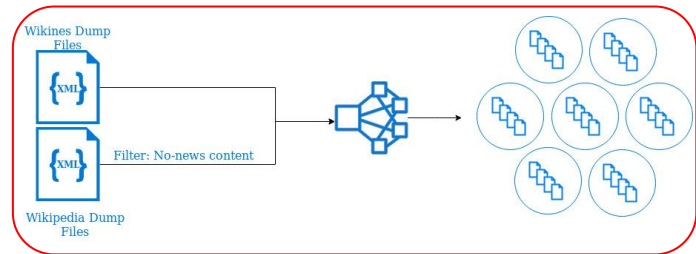
Dataset: What do we need?

Neural Abstractive:

- Typically needs a dataset of thousands of documents
- i.e. CNN/Dailymail > 90k/197k (single-document dataset)

Dataset: What do we have?

- Multi-Document datasets are typically small
- One of the most well-known does not contain more than 60 cluster and 600 documents

Data Source	Cluster/Sample Summaries	Documents
DUC 2001	30	309
DUC 2002	59	567
DUC 2004	50	500
Total	**139**	**1,376**

Dataset: Solution

- We created Webis-wikinews-corpus
- One of the first of its kind...
  - Large-scale
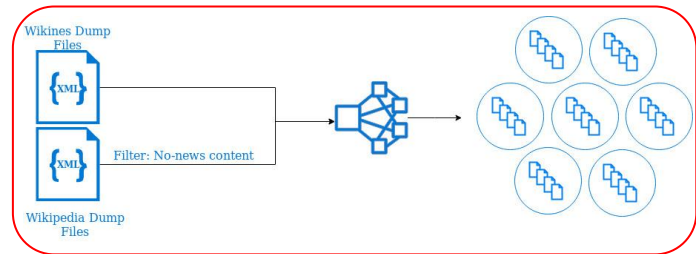 - Multi-document
  - For the news domain

Dataset: Source

- Wikimedia Projects : Wikinews & Wikipedia
    - Unbiased
    - Open-source
    - Up-to-date
    - Clustered news from reliable sources

Dataset: Construction

Extract the useful information from Dump File:

- Article, source links, auxiliary information
- Only the pages with news sources for the Wikipedia

Dataset: Construction

Retrieval:

Page Request → 404 → NO → Download → Extract → Filter →

YES → Wayback Machine
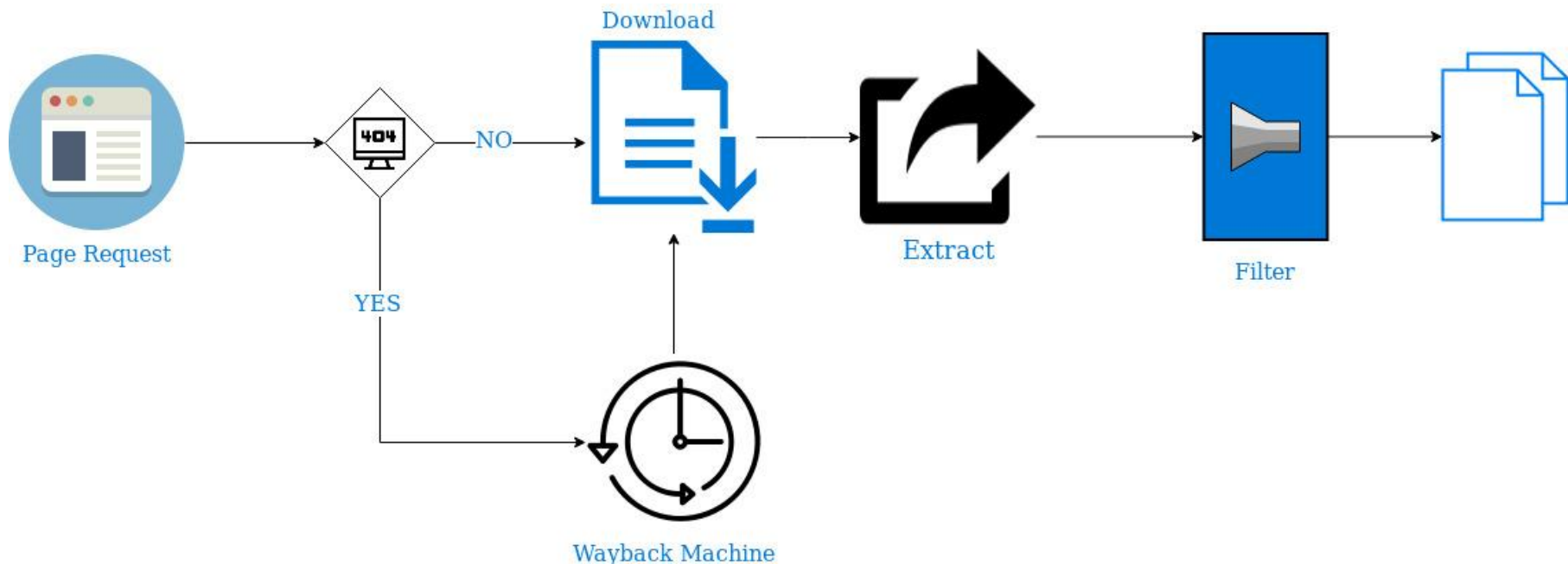
Dataset: Size & Folder Structure

Data Source	Cluster/Sample Summaries	Documents
Wikinews	9,514	21,314
Wikipedia	2,174	17,807
Total	**11,688**	**39,121**
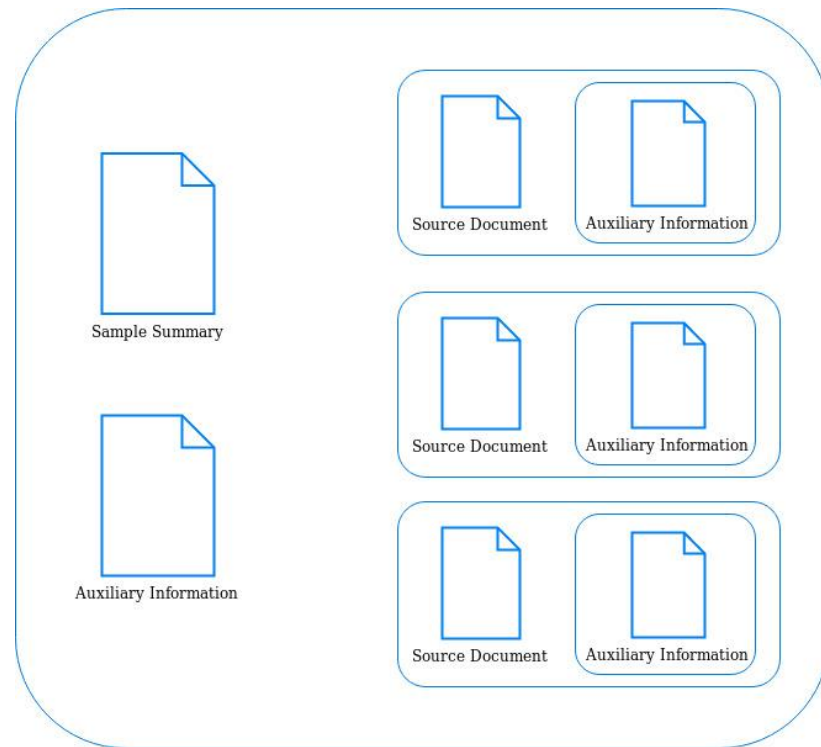
Unified Summarization Pipeline
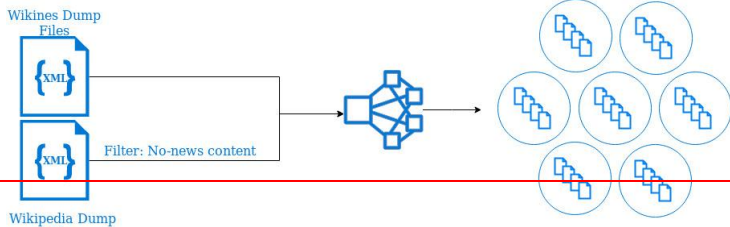
Wikines Dump Files

{XML}

Filter: No-news content

{XML}

Wikipedia Dump Files

Wikisummarizer

Pointer-Generator Network

Extractive Summary

Abstractive Summary

Target

Training Models

Evaluation Methods

double-abstractive   ea-full-target   ea-short-target   Automatic Evaluation   Manual Evaluation   Observations
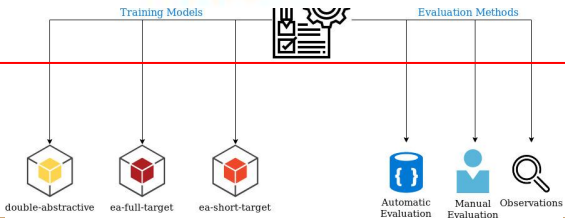
Unified Summarization

- Extractive Summarization: Wikisummarizer

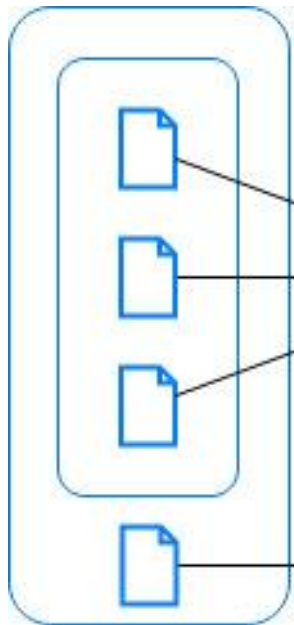- Abstractive Summarization: Pointer-Generator Network [See et al., 2017]
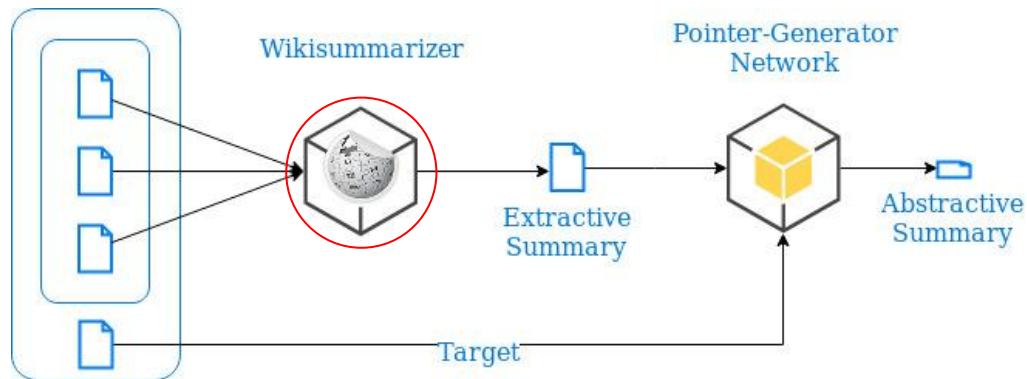
Unified Summarization

- Extractive Summarization: Wikisummarizer
    - A Google Brain project [Liu et al. ,2018] : Extraction from similar source (Wikipedia)

- Abstractive Summarization: Pointer-Generator Network [See et al., 2017]

Unified Summarization

- Extractive Summarization: Wikisummarizer
  - A Google Brain project [Liu et al. ,2018] : Extraction from similar source (Wikipedia)
  - CST: Filter out the duplication [Radev and Zhang, 2004]
- Abstractive Summarization: Pointer-Generator Network [See et al., 2017]

Unified Summarization

- Extractive Summarization: Wikisummarizer
  - A Google Brain project [Liu et al. ,2018] : Extraction from similar source (Wikipedia)
  - CST: Filter out the duplication [Radev and Zhang, 2004]
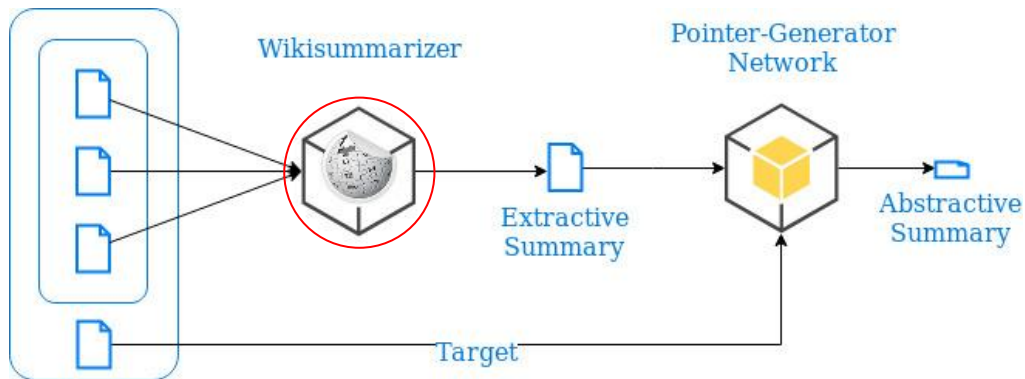- Abstractive Summarization: Pointer-Generator Network [See et al., 2017]

Unified Summarization

- Extractive Summarization: Wikisummarizer
  - A Google Brain project [Liu et al. ,2018] : Extraction from similar source (Wikipedia)
  - CST: Filter out the duplication [Radev and Zhang, 2004]
- Abstractive Summarization: Pointer-Generator Network [See et al., 2017]
  - Solves the problems of earlier approaches such as repetitiveness, senseless sentences and inaccurate facts
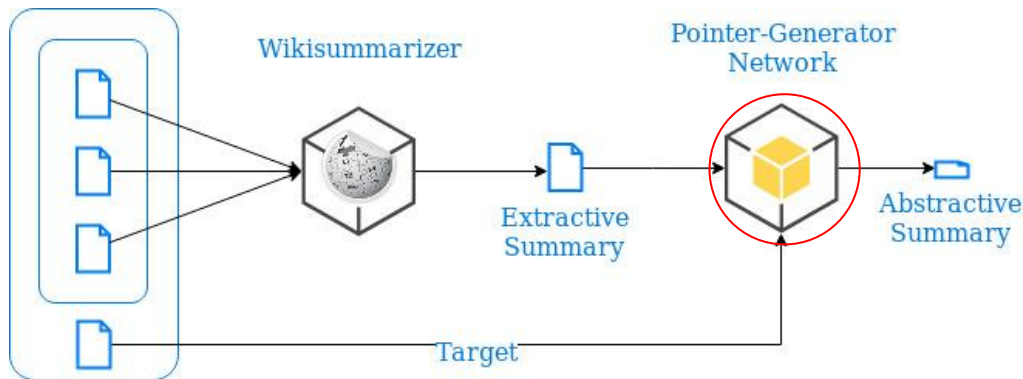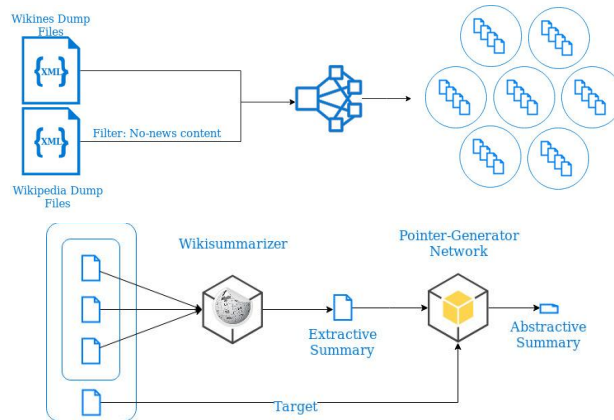
Experiments&Evaluation

Experiment

Training Models

Evaluation Methods

double-abstractive   ea-full-target   ea-short-target

Automatic Evaluation   Manual Evaluation   Observations
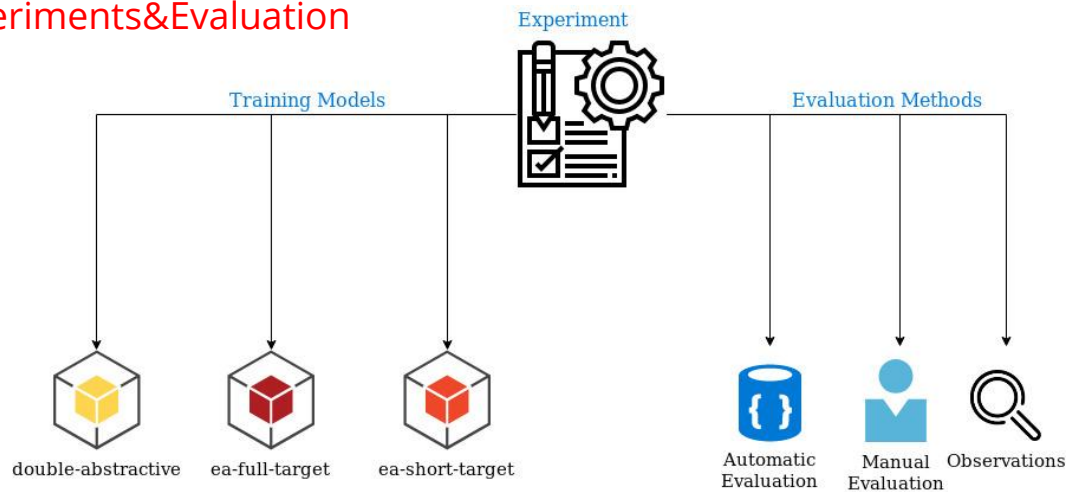
Experiments and Evaluation: Training Models

- Double-abstractive
- Extractive + Abstractive Full Target
- Extractive + Abstractive Short Target
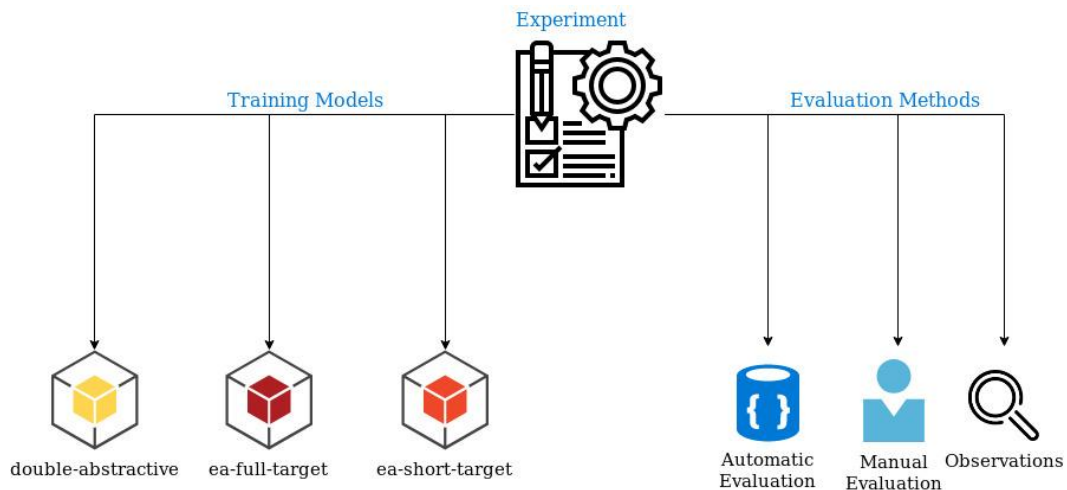
Experiments and Evaluation: Training Models

Double-abstractive
- Trivial method
- To examine the unified model
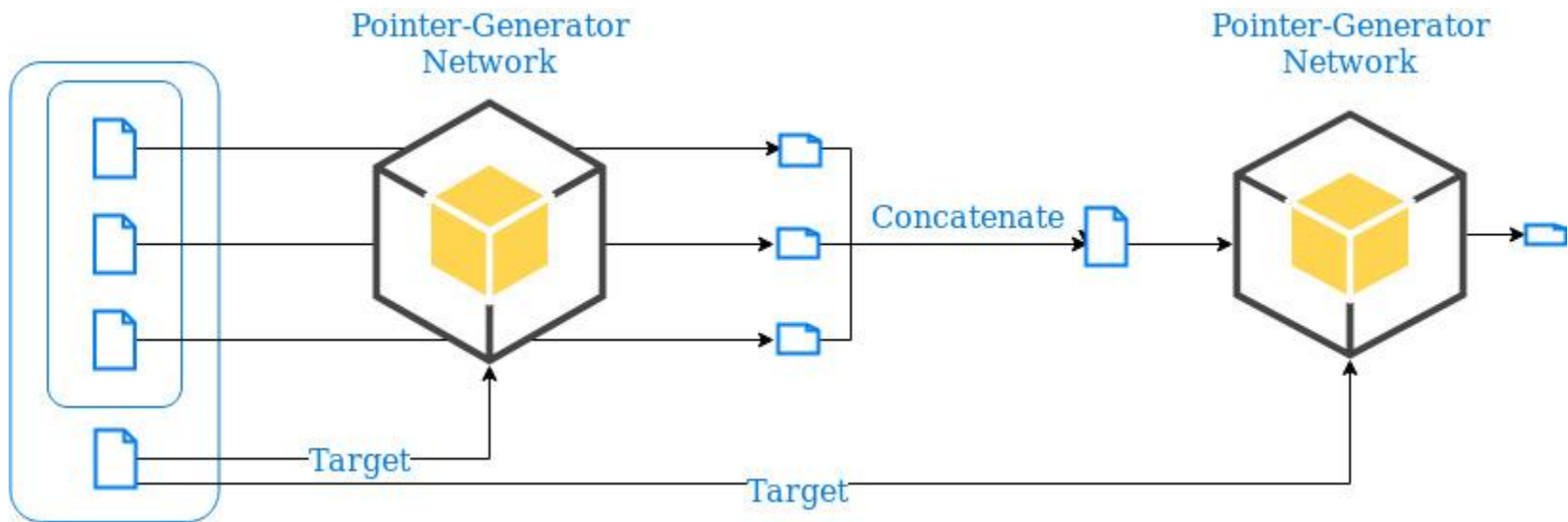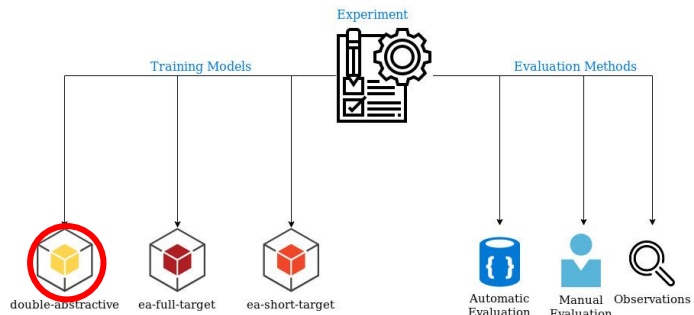
Experiments and Evaluation: Training Models

Unified Models: Extractive + Abstractive
- ea-full-target - Target document size : Full size
- ea-short-target - Target document size : 3 sentences
- To examine the effects of different ratio between input and target
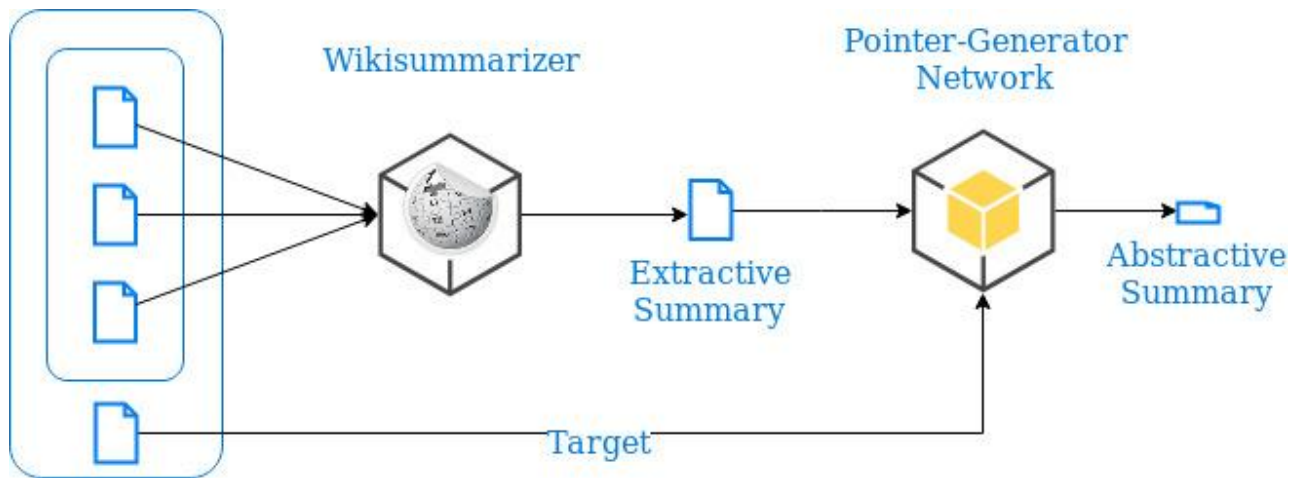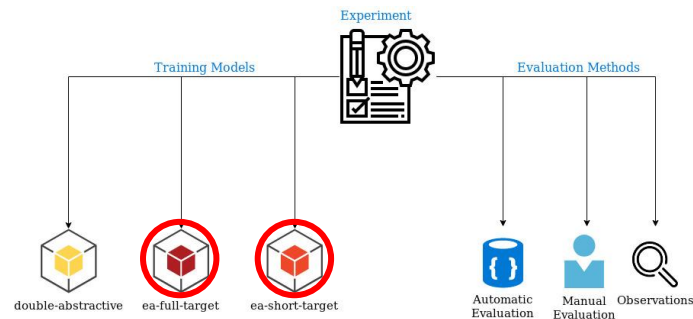
Introduction : Automatic Summarization for Journalism

"Journalism is the activity of *gathering, assessing, creating, and presenting* *news and information."*

Experiments and Evaluation: Aspects

"Journalism is the activity of *gathering, assessing,* *creating, and presenting* *news and information."*

- Aspects :
  - Content
  - Readability

Experiments and Evaluation: Aspects

- Aspects :
  - Content
    - Automatic > a state-of-the-art method exist
  - Readability
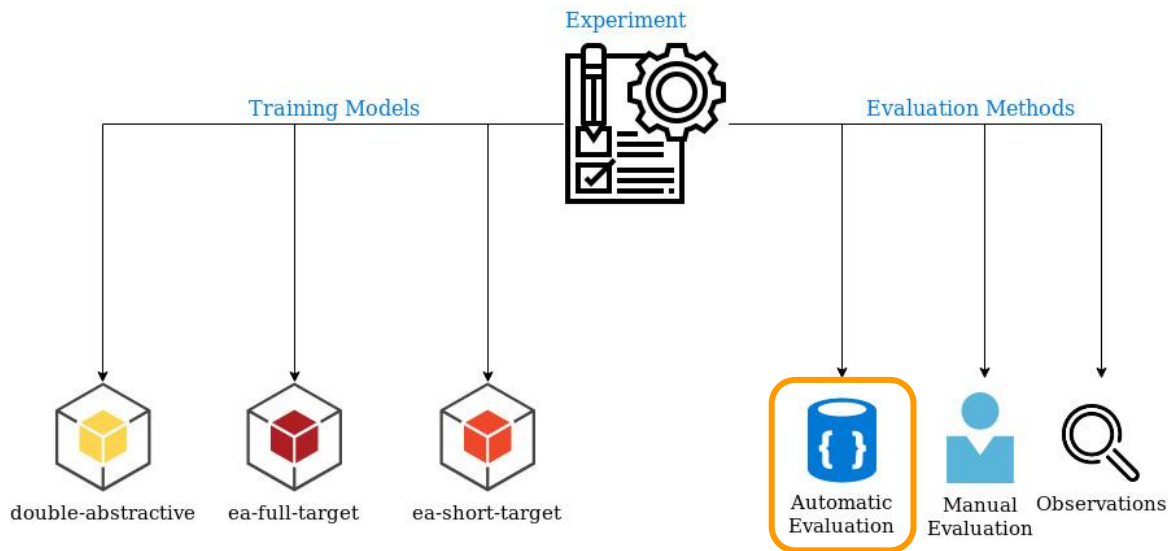
Experiments and Evaluation: ROUGE

Computer Generated Summary : <mark>the cat was</mark> found <mark>under the bed</mark>

Ground-truth Summary : <mark>the cat was under the bed</mark>

$$Recall = \frac{number\_of\_overlapping\_text\_units}{total\_words\_in\_ground\_truth} = \frac{6}{6} = 1$$

Experiments and Evaluation: ROUGE

Computer Generated Summary : ==the cat was== found ==under the bed==

Ground-truth Summary : ==the cat was under the bed==

Experiments and Evaluation: ROUGE

Computer Generated Summary : <mark>the cat was</mark> found <mark>under the bed</mark>

Ground-truth Summary : <mark>the cat was under the bed</mark>

$$Precision = \frac{number\_of\_overlapping\_text\_units}{total\_words\_in\_generated\_text} = \frac{6}{7} = 0.86$$

Experiments and Evaluation: ROUGE

Computer Generated Summary : <mark>the cat was</mark> found <mark>under the bed</mark>

Ground-truth Summary : <mark>the cat was under the bed</mark>

$$F - Measure = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

Experiments and Evaluation: ROUGE

- ROUGE-N(ROUGE-1) : Overlapping n-grams > Word wise similarity
- ROUGE-L : Longest Common Subsequence > Sequence wise similarity

Experiments and Evaluation: Results

- Aspects :
  - Content:
    - Automatic > a state-of-the-art method exist

ROUGE	double-abstractive	ea-full-target
ROUGE-1	0.23	0.29
ROUGE-L	0.16	0.21

Experiments and Evaluation: Results

- Aspects :
    - Content
        - Automatic > a state-of-the-art method exist

ROUGE	double-abstractive	ea-full-target	ea-short-target
ROUGE-1	0.23	0.29	0.54
ROUGE-L	0.16	0.21	0.49

Experiments and Evaluation: Aspects

- Aspects :
  - Content
    - Automatic > a state-of-the-art method exist
  - Readability
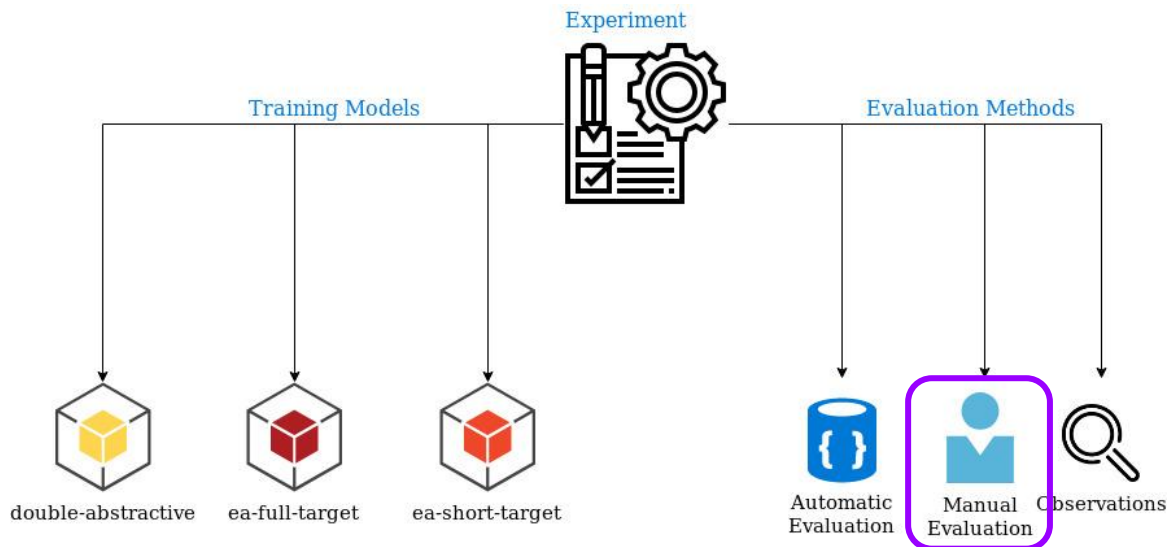
Experiments and Evaluation: ROUGE for readability?

Computer Generated Summary : was the found under the cat

Ground-truth Summary : the cat was found under the bed

1 ROUGE-1 Average_R: 0.83333
1 ROUGE-1 Average_P: 0.83333
1 ROUGE-1 Average_F: 0.83333

1 ROUGE-L Average_R: 0.50000
1 ROUGE-L Average_P: 0.50000
1 ROUGE-L Average_F: 0.50000

Experiments and Evaluation: ROUGE for readability?

Computer Generated Summary : ==was the found under the cat==

Ground-truth Summary : ==the cat was found under the== bed

1 ROUGE-1 Average_R: 0.83333
1 ROUGE-1 Average_P: 0.83333
1 ROUGE-1 Average_F: 0.83333

1 ROUGE-L Average_R: 0.50000
1 ROUGE-L Average_P: 0.50000
1 ROUGE-L Average_F: 0.50000

Computer Generated Summary : ==he== found no lights on

Ground-truth Summary : all of the lamps were off already when ==he== walked into the room

1 ROUGE-1 Average_R: 0.07692
1 ROUGE-1 Average_P: 0.20000
1 ROUGE-1 Average_F: 0.11111

1 ROUGE-L Average_R: 0.07692
1 ROUGE-L Average_P: 0.20000
1 ROUGE-L Average_F: 0.11111

Experiments and Evaluation: Aspects

- Aspects :
  - Content
    - Automatic > a state-of-the-art method exist
  - Readability
    - ROUGE is not reliable for readability
    - Manual > There are not many automatic methods, mostly manual
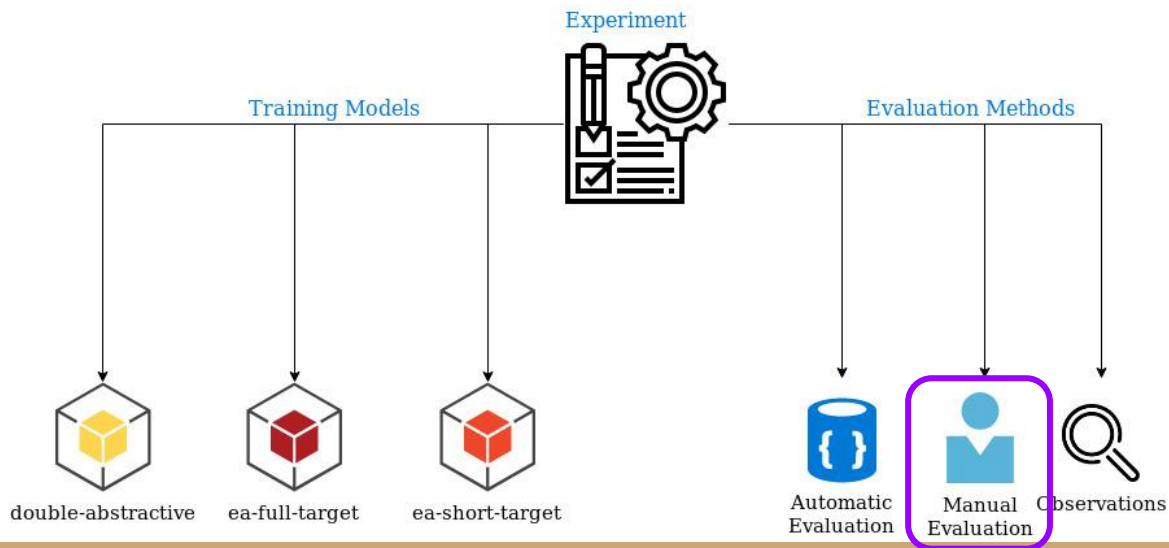
Experiments and Evaluation: Readability Aspects by DUC

- Grammaticality
- Non-redundancy
- Referential clarity
- Focus
- Structure and coherence

Experiments and Evaluation: Survey

- Grammaticality
- Non-redundancy
- Referential clarity
- Focus
- Structure and coherence

First Survey

Experiments and Evaluation: Survey

- Gramm
- Non-re
- Refere
- Focus
- Struct

First Surv
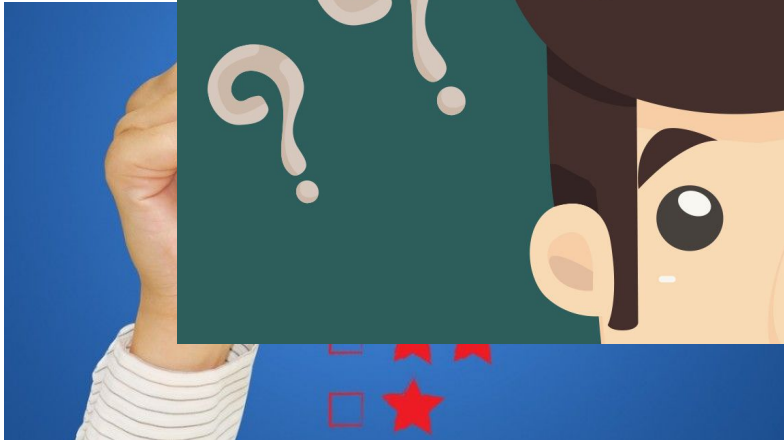
Experiments and Evaluation: Survey

- Grammaticality
- Non-redundancy
- Referential clarity
- Focus
- Structure and coherence

First Survey

Second Survey

Experiments and Evaluation: Results

- Aspects :
  - Content:
    - Automatic > a state-of-the-art method exist
  - Readability
    - ROUGE is not reliable for readability
    - Manual > There are not many automatic methods, mostly manual

Training Model	Mean Score
double-abstractive	2.15
ea-full-target	2.67

Experiments and Evaluation: Results

- Aspects :
  - Content:
    - Automatic > a state-of-the-art method exist
  - Readability
    - ROUGE is not reliable for readability
    - Manual > There are not many automatic methods, mostly manual

Training Model	Mean Score
double-abstractive	2.15
ea-full-target	2.67
ea-short-target	4.18

Experiments and Evaluation: Comparison of Evaluation Aspects
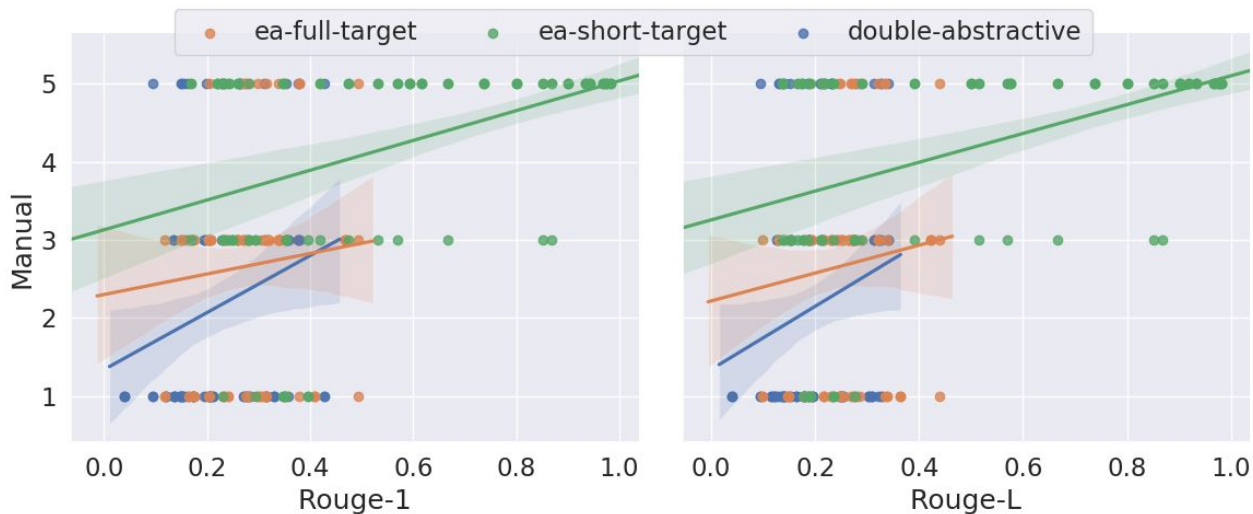
Training Model	Mean Score
double-abstractive	2.15
ea-full-target	2.67
ea-short-target	4.18

ROUGE	double-abstractive	ea-full-target	ea-short-target
ROUGE-1	0.23	0.29	0.54
ROUGE-L	0.16	0.21	0.49

Observations

Experiment

Training Models

Evaluation Methods

double-abstractive   ea-full-target   ea-short-target

Automatic
Evaluation

Manual
Evaluation
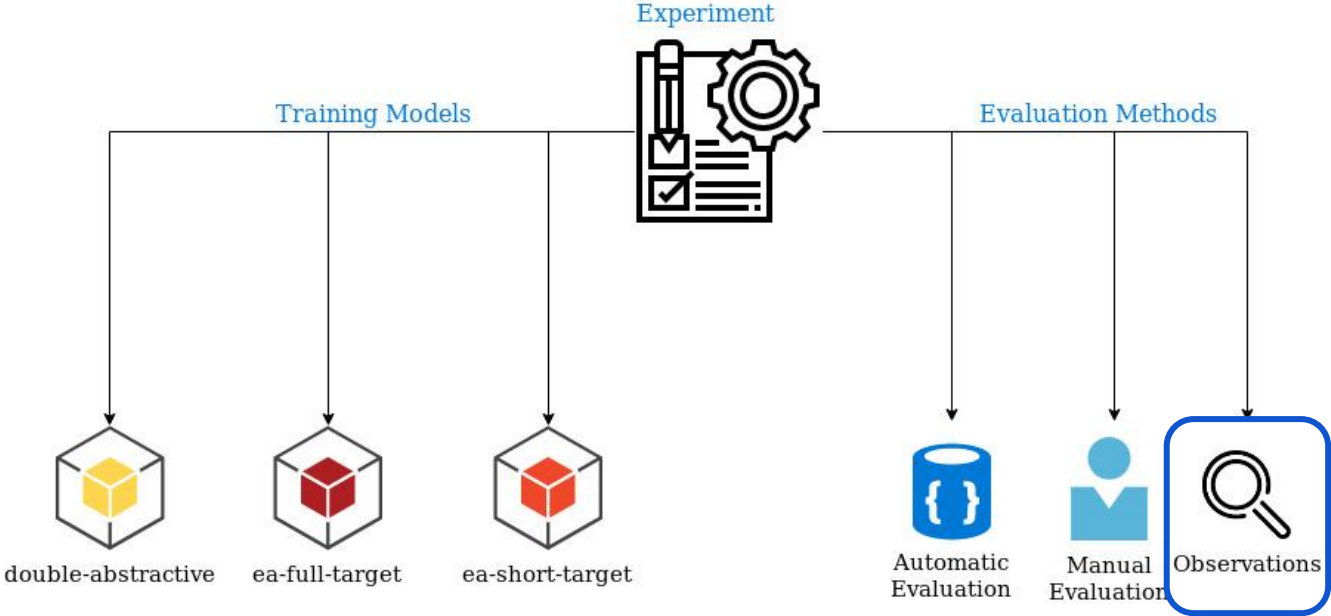
Observations

Observations : UNK Error

Ground-truth Summary: yesterday san francisco giants lf barry bonds <mark>hit a 435-foot home run</mark> , his 756th , off a pitch from <mark>mike bacsick of the</mark> washington nationals , breaking the all-time career home run record , formerly held by hank aaron.the pitch , the seventh of the at-bat , was a 3-2 pitch , which bonds hit into the right-center field bleachers.matt murphy , a 22-year-old from queens in new york city , got the ball and was promptly protected and escorted away from the mayhem by a group of san francisco police officers .

Computer Generated Summary: yesterday san francisco giants [UNK] barry bonds <mark>hit a [UNK]home run</mark> , his 756th , off a pitch from <mark>mike [UNK] of the</mark> washington nationals , breaking the [UNK] home run in 1974 .

Observations : Inaccurate Facts

Input
…
Hello Kitty was first introduced by ==Japanese company Sanrio in 1974.==The cute round-faced
…

Ground-truth summary: the armband , which features "hello kitty" sitting on top of two hearts ,
will be worn by police officers who commit minor offences.these include , and parking in a prohibited
area.the officers will also be forced to stay with ==the deputy chief all day in division office== and
will be forbidden to disclose their offences.

Computer-generated summary: the armband , which features "hello kitty" sitting on top of
two hearts , will be worn by police officers who commit minor [UNK] include , and parking in a
prohibited area.the officers will also be forced to stay with ==the deputy police in 1974.==

Observations : Repetition

Computer-generated summary: a court in , kenya has sentenced a group of seven somali pirates to five years each in jail , according to a statement by the european [UNK] mission eu [UNK] said that the men , " i have concrete proof that you attacked a vessel in the high seas and i order you to serve five years in prison seas and i order you to serve five years in prison seas and i order you to serve five years in prison seas and i order you to serve five years in prison seas and i order you to serve

Observations : Successful Summaries

Ground-truth summary: At least ten attackers with knives, dressed in black, attacked a train station in , China yesterday.At least 28 victims were killed, with 113 more wounded by knives, Chinese state news agency reported.The local municipal government accuses " separatist forces" for the attack.

Computer-generated summary: At least ten attackers with knives, dressed in black, attacked a train station in , China yesterday.At least 28 victims were killed, with 113 more wounded by knives,Chinese state news agency reported.The local municipal government accuses " separatist forces" for the attack.

Ground-truth summary: the united states navy has successfully destroyed a crippled spy satellite in a decaying orbit , by intercepting it with a missile.a modified sm-3 missile was launched from the uss lake erie at 03:26 gmt this morning , and intercepted the usa-193 satellite around three minutes later.it has been reported that the satellite has broken into around 80 pieces , some of which have already re-entered the earth 's atmosphere .

Computer-generated summary: the united states navy has successfully destroyed a crippled spy satellite in a decaying orbit , by intercepting it with a missile.a modified sm-3 missile was launched from the uss lake erie at 03:26 gmt this morning , and intercepted the usa-193 satellite around three minutes later.it has been reported that the satellite has been damaged.

Recap, Conclusion & Future Work

- Contributions
    - One of the first large-scale multi-document summarization dataset for news domain

Recap, Conclusion & Future Work

- Contributions
  - One of the first large-scale multi-document summarization dataset for news domain
  - Wikisummarizer

Recap, Conclusion & Future Work

- Contributions
  - One of the first large-scale multi-document summarization dataset for news domain
  - Wikisummarizer
  - Unified multi-document summarization pipeline

Recap, Conclusion & Future Work

- Contributions
    - One of the first large-scale multi-document summarization dataset for news domain
    - Wikisummarizer
    - Unified multi-document summarization pipeline
- Conclusion

Recap, Conclusion & Future Work

- Contributions
  - One of the first large-scale multi-document summarization dataset for news domain
  - Wikisummarizer
  - Unified multi-document summarization pipeline
- Conclusion
  - Extractive Summarization proved to be a good method to transfer from multi-document to single-document

Recap, Conclusion & Future Work

- Contributions
  - One of the first large-scale multi-document summarization dataset for news domain
  - Wikisummarizer
  - Unified multi-document summarization pipeline
- Conclusion
  - Extractive Summarization proved to be a good method to transfer from multi-document to single-document
  - Better content selection resulted in better readability

Recap, Conclusion & Future Work

- Contributions
  - One of the first large-scale multi-document summarization dataset for news domain
  - Wikisummarizer
  - Unified multi-document summarization pipeline
- Conclusion
  - Extractive Summarization proved to be a good method to transfer from multi-document to single-document
  - Better content selection resulted in better readability
  - Even though there is a room for improvement, the idea behind the framework is promising

Recap, Conclusion & Future Work

- Contributions
  - One of the first large-scale multi-document summarization dataset for news domain
  - Wikisummarizer
  - Unified multi-document summarization pipeline
- Conclusion
  - Extractive Summarization proved to be a good method to transfer from multi-document to single-document
  - Better content selection resulted in better readability
  - Even though there is a room for improvement, the idea behind the framework is promising
- Future Work

Recap, Conclusion & Future Work

- Contributions
  - One of the first large-scale multi-document summarization dataset for news domain
  - Wikisummarizer
  - Unified multi-document summarization pipeline
- Conclusion
  - Extractive Summarization proved to be a good method to transfer from multi-document to single-document
  - Better content selection resulted in better readability
  - Even though there is a room for improvement, the idea behind the framework is promising
- Future Work
  - Extending the dataset

Recap, Conclusion & Future Work

- Contributions
  - One of the first large-scale multi-document summarization dataset for news domain
  - Wikisummarizer
  - Unified multi-document summarization pipeline
- Conclusion
  - Extractive Summarization proved to be a good method to transfer from multi-document to single-document
  - Better content selection resulted in better readability
  - Even though there is a room for improvement, the idea behind the framework is promising
- Future Work
  - Extending the dataset
    - Distant Supervision for clustering other datasets

Recap, Conclusion & Future Work

- Contributions
  - One of the first large-scale multi-document summarization dataset for news domain
  - Wikisummarizer
  - Unified multi-document summarization pipeline
- Conclusion
  - Extractive Summarization proved to be a good method to transfer from multi-document to single-document
  - Better content selection resulted in better readability
  - Even though there is a room for improvement, the idea behind the framework is promising
- Future Work
  - Extending the dataset
    - Distant Supervision for clustering other datasets
    - Merge with Multi-News

Recap, Conclusion & Future Work

- Contributions
  - One of the first large-scale multi-document summarization dataset for news domain
  - Wikisummarizer
  - Unified multi-document summarization pipeline
- Conclusion
  - Extractive Summarization proved to be a good method to transfer from multi-document to single-document
  - Better content selection resulted in better readability
  - Even though there is a room for improvement, the idea behind the framework is promising
- Future Work
  - Extending the dataset
    - Distant Supervision for clustering other datasets
    - Merge with Multi-News
  - Other setups of Wikisummarizer

Thanks for coming...

References

[Liu et al. ,2018] : Peter J. Liu, Mohammad Saleh, Etienne Pot, Ben Goodrich, Ryan Sepassi,Lukasz Kaiser, and Noam Shazeer. Generating wikipedia by summarizing long sequences. arXiv:1801.10198 [cs], 2018. URL http://arxiv.org/abs/1801.10198

[Radev and Zhang, 2004] : Dragomir R. Radev and Zhu Zhang. Cross-document relationship classification for text summarization. 2004.

[See et al., 2017] : Abigail See, Peter J. Liu, and Christopher D. Manning. Get to the point: Summarization with pointer-generator networks. In ACL, 2017.

Image references

https://img.buzzfeed.com/buzzfeed-static/static/enhanced/webdr05/2013/6/28/10/enhanced-buzz-29020-1372431014-2.jpg?downsize=800:*&output-format=auto&output-quality=auto

https://tr.pinterest.com/pin/451345193878807678/?lp=true

https://images.sadhguru.org/sites/default/files/media_files/iso/en/48257-confusion-clarity-spiritual-path.jpg

https://www.timeshighereducation.com/sites/default/files/styles/the_breaking_news_image_style/public/Pictures/web/n/c/o/numbers_on_podium.jpg?itok=-nVlhkPx