

CSCE 590-1: From Data to Decisions with Open Data: A Practical Introduction to AI

Lecture 22: Project Report, Summarization

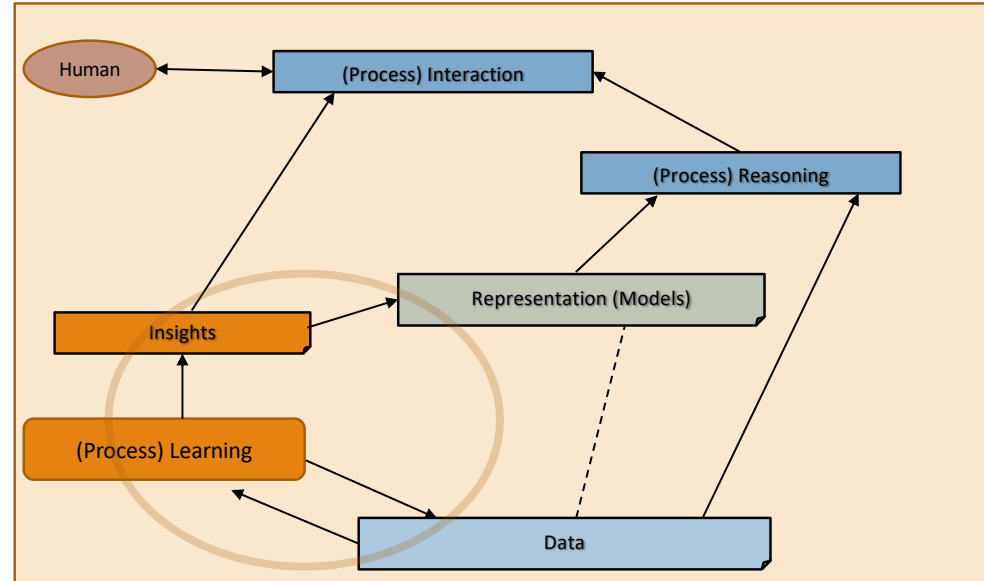
PROF. BIPLAV SRIVASTAVA, AI INSTITUTE

1ST APRIL, 2021

Carolinian Creed: “I will practice personal and academic integrity.”

Organization of Lecture 22

- Introduction Segment
 - Recap/ Discussion of Lecture 21
- Main Segment
 - Project Review - Interim
 - Summarization
 - Sentiment Analysis
- Concluding Segment
 - About Next Lecture – Lecture 23
 - Ask me anything



Introduction Segment

Upcoming Classes

21	Mar 25 (Th)	Review: project presentations, Discussion	
22	Mar 30 (Tu)	Wellness Holiday	
23	Apr 1 (Th)	Text: Summary	
24	Apr 6 (Tu)	Text: Representation, Sentiment	
25	Apr 8 (Th)	Text: Sentiment, Visualization	
26	Apr 13 (Tu)	Case Study: Finance – Guest lecture	
27	Apr 15 (Th)	Paper presentations – Graduate students	Quiz 4 // Final assignment for Graduate students
28	Apr 20 (Tu)	Advanced: Bias and Trust Issues	
30	Apr 22 (Th)	Project presentations	
31	Apr 27 (Tu)	Reading day	



Paper Presentations – Graduate Students

- Select a paper appearing at a top-AI or data conference (AAAI, IJCAI, NeurIPS, SIGMOD, WWW, ICML, VLDB, ...) during 2019-2021
- Present in class for 10 + 5 minutes of Q/A
- Things to cover
 - **Summary:** problem, solution, related work, contributions
 - **Opinion:** What you liked or did not like
- Please enter by end of today

Recap of Lecture 21

Review of student projects

Completed

10	Brendan	Curran
9	Sarayu	Das
8	Marilyn	Gartley
7	Vedant	Khandelwal
6	Vishal	Pallagani
5	Avineet Kumar	Singh
4	Ahad Hasan	Tanim
3	James	Thompson
2	Rohit	Naini
1	Terric	Taylor

Main Segment

Common NLP Tasks

- **Summarization**
- **Sentiment**
- Machine translation
- Machine Comprehension
- Natural Language Interface to Databases
- Natural Language Generation

Motivation for Summarization

- We have a variety of data around
 - Volume: Single document, Multiple documents
 - Variety: Text, numbers/ tables, figures, ...
 - Languages
- Can one get a quick insight about the data ? Meaning ?
 - Choice 1: **Important** ideas in the input
 - Choice 2: **Representative** ideas in the input
 - Choice 3: **Example** ideas in the input
 - Random selection is sufficient?

Example Text

1. <https://www.nytimes.com/article/2020-nobel-prize-winners.html>
2. 2020 Nobel Prize Winners: Full List
3. Nobel Prize season begins every October as committees in Sweden and Norway name laureates in a variety of prizes in the sciences, literature and economics, as well as peace work. The announcements started last week with the awarding of the prize in Physiology or Medicine. They wrapped up on Monday, when the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel was announced.
4. The Nobel Prizes most years are presented to recipients in Stockholm and Oslo in December. Because of the coronavirus pandemic, the committees are changing their approaches. Some of the events in Stockholm will be canceled in favor of a digital ceremony for the Nobelists, and medals and diplomas are to be distributed to the recipients' embassies and handed over in their home countries. Recipients may be invited to the award ceremony for 2021, if possible.
5. The Oslo ceremony for the peace prize will be smaller than in most years, with a limited audience.
6. The Nobel committee also announced another change last month: Each prize will rise to 10 million Swedish krona, 1 million more than in the previous year. That's a hike in the prize value of about \$112,000 in current exchange rates.
7. The 2020 Winners
8. Physiology or Medicine
9. Drs. Harvey J. Alter, Michael Houghton and Charles M. Rice on Monday received the prize for their discovery of the hepatitis C virus. The Nobel committee said the three scientists had "made possible blood tests and new medicines that have saved millions of lives."
10. Physics
11. Roger Penrose, Reinhard Genzel and Andrea Ghez are the recipients of the Nobel Prize in Physics for 2020. Credit...Pool photo by Fredrik Sandberg
12. Roger Penrose, Reinhard Genzel and Andrea Ghez received the prize on Tuesday for their discoveries that have improved understanding of the universe, including work on black holes.
13. Chemistry
14. The Nobel Prize in Chemistry was jointly awarded on Wednesday to Emmanuelle Charpentier and Jennifer A. Doudna for their work on the development of Crispr-Cas9, a method for genome editing.
15. Literature
16. The Nobel Prize in Literature was awarded on Thursday to Louise Glück, one of America's most celebrated poets, "for her unmistakable poetic voice that with austere beauty makes individual existence universal."
17. Peace Prize
18. The Nobel Peace Prize was awarded on Friday to the World Food Program for its efforts to combat a surge in global hunger amid the coronavirus pandemic, which has swept around the world with devastating impact.
19. Economic Science
20. Paul R. Milgrom and Robert B. Wilson were awarded the Nobel in economic science on Monday for improvements to auction theory and inventions of new auction formats.

A Few Ways to Generate Insights

- Word tag cloud (textual data; based on frequency)
 - Important
- Topic analysis (textual data, based on learned patterns)
 - Important
- **Summary generation** – formal study
 - Representative
 - Important
 - Example

Example Text

Summary Generated

1. 2020 Nobel Prize Winners: Full List Nobel Prize season begins every October as committees in Sweden and Norway name laureates in a variety of prizes in the sciences, literature and economics, as well as peace work. ?
2. Peace Prize The Nobel Peace Prize was awarded on Friday to the World Food Program for its efforts to combat a surge in global hunger amid the coronavirus pandemic, which has swept around the world with devastating impact.
3. Literature The Nobel Prize in Literature was awarded on Thursday to Louise Glück, one of America's most celebrated poets, ?for her unmistakable poetic voice that with austere beauty makes individual existence universal.

1. <https://www.nytimes.com/article/2020-nobel-prize-winners.html>
2. 2020 Nobel Prize Winners: Full List
3. Nobel Prize season begins every October as committees in Sweden and Norway name laureates in a variety of prizes in the sciences, literature and economics, as well as peace work. The announcements started last week with the awarding of the prize in Physiology or Medicine. They wrapped up on Monday, when the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel was announced.
4. The Nobel Prizes most years are presented to recipients in Stockholm and Oslo in December. Because of the coronavirus pandemic, the committees are changing their approaches. Some of the events in Stockholm will be canceled in favor of a digital ceremony for the Nobelists, and medals and diplomas are to be distributed to the recipients' embassies and handed over in their home countries. Recipients may be invited to the award ceremony for 2021, if possible. The Oslo ceremony for the peace prize will be smaller than in most years, with a limited audience.
5. The Nobel committee also announced another change last month: Each prize will rise to 10 million Swedish krona, 1 million more than in the previous year. That's a hike in the prize value of about \$112,000 in current exchange rates.
7. The 2020 Winners
8. Physiology or Medicine
9. Drs. Harvey J. Alter, Michael Houghton and Charles M. Rice on Monday received the prize for their discovery of the hepatitis C virus. The Nobel committee said the three scientists had ?made possible blood tests and new medicines that have saved millions of lives.?
10. Physics
11. Roger Penrose, Reinhard Genzel and Andrea Ghez are the recipients of the Nobel Prize in Physics for 2020. Credit... Pool photo by Fredrik Sandberg
12. Roger Penrose, Reinhard Genzel and Andrea Ghez received the prize on Tuesday for their discoveries that have improved understanding of the universe, including work on black holes.
13. Chemistry
14. The Nobel Prize in Chemistry was jointly awarded on Wednesday to Emmanuelle Charpentier and Jennifer A. Doudna for their work on the development of Crispr-Cas9, a method for genome editing.
15. Literature
16. The Nobel Prize in Literature was awarded on Thursday to Louise Glück, one of America's most celebrated poets, ?for her unmistakable poetic voice that with austere beauty makes individual existence universal.?
17. Peace Prize
18. The Nobel Peace Prize was awarded on Friday to the World Food Program for its efforts to combat a surge in global hunger amid the coronavirus pandemic, which has swept around the world with devastating impact.
19. Economic Science
20. Paul R. Milgrom and Robert B. Wilson were awarded the Nobel in economic science on Monday for improvements to auction theory and inventions of new auction formats.

Summarization Type and Methods

- **Extractive:** extract summary **with** content in the input
 - Should be able to trace output back to the input
 - Interpretability is easy to show
- **Abstractive:** create summary **from ideas** in the content in the input
 - but using at least some different words
 - Useful in creative task – digital journalism
- **Compressive:** remove redundant from input
 - shorter sentence by removing redundant information
 - Preserve grammar and important content in the input

Evaluation of Text Summarization

- **Intrinsic** – How good is the output ?
- **Considerations**
 - Quality of output - readability
 - Coherence of information – out of context sentences ?
 - Compression of information

Comparison against a reference summary

- Is this the only summary possible?
- A scale of what summary should contain
- Constraints on summary characteristics.
E.g., number of words

Where is the reference summary?

- **Experts do not agree on summaries**
 - Variance between experts
 - Not even their when separated 8-weeks apart nearly 50% times
- **But experts often agree on important sentences** that make up into a summary

Summarization Evaluation: An Overview, Inderjeet Mani,
<http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings2/sum-mani.pdf> 2019

Evaluation of Text Summarization

- **Intrinsic** – How good is the output ?
- **Considerations**
 - Quality of output - readability
 - Coherence of information – out of context sentences ?
 - Compression of information

Automatic Evaluation

- Against a set of criteria
- Criteria can be task independent or task-dependent
- Can also be about sentences to include and their rank
- Can also be about topics in content and in summary

Metrics

- ROUGE (Recall Oriented Understudy for Gisting Evaluation) score
- METEOR (Metric for Evaluation of Translation with Explicit Ordering) score
- BLEU (bilingual evaluation understudy) score
- SUMMAC summary automatic scoring

Summarization Evaluation: An Overview, Inderjeet Mani,
<http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings2/sum-mani.pdf> 2019

Evaluation of Text Summarization

- **Extrinsic** – What is the impact of the output ?
- Overall idea
 - Describe a task that a person wants to do with a summary
 - Given the generated summary, how well can a person do that job
- Comments
 - Extrinsic useful for business sponsor but not directly to technical developer

Summarization Evaluation: An Overview, Inderjeet Mani,
<http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings2/sum-mani.pdf> 2019

Coding Illustration - Extractive

<https://github.com/biplav-s/course-nl/tree/master/l20-textsumm>

Evaluation in Example Text

Summary Generated

1. 2020 Nobel Prize Winners: Full List Nobel Prize season begins every October as committees in Sweden and Norway name laureates in a variety of prizes in the sciences, literature and economics, as well as peace work. ?
2. Peace Prize The Nobel Peace Prize was awarded on Friday to the World Food Program for its efforts to combat a surge in global hunger amid the coronavirus pandemic, which has swept around the world with devastating impact.
3. Literature The Nobel Prize in Literature was awarded on Thursday to Louise Glück, one of America's most celebrated poets, ?for her unmistakable poetic voice that with austere beauty makes individual existence universal.

1. <https://www.nytimes.com/article/2020-nobel-prize-winners.html>
2. 2020 Nobel Prize Winners: Full List
3. Nobel Prize season begins every October as committees in Sweden and Norway name laureates in a variety of prizes in the sciences, literature and economics, as well as peace work. The announcements started last week with the awarding of the prize in Physiology or Medicine. They wrapped up on Monday, when the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel was announced.
4. The Nobel Prizes most years are presented to recipients in Stockholm and Oslo in December. Because of the coronavirus pandemic, the committees are changing their approaches. Some of the events in Stockholm will be canceled in favor of a digital ceremony for the Nobelists, and medals and diplomas are to be distributed to the recipients' embassies and handed over in their home countries. Recipients may be invited to the award ceremony for 2021, if possible. The Oslo ceremony for the peace prize will be smaller than in most years, with a limited audience.
6. The Nobel committee also announced another change last month: Each prize will rise to 10 million Swedish krona, 1 million more than in the previous year. That's a hike in the prize value of about \$112,000 in current exchange rates.
7. The 2020 Winners
8. Physiology or Medicine
9. Drs. Harvey J. Alter, Michael Houghton and Charles M. Rice on Monday received the prize for their discovery of the hepatitis C virus. The Nobel committee said the three scientists had ?made possible blood tests and new medicines that have saved millions of lives.?
10. Physics
11. Roger Penrose, Reinhard Genzel and Andrea Ghez are the recipients of the Nobel Prize in Physics for 2020. Credit... Pool photo by Fredrik Sandberg
12. Roger Penrose, Reinhard Genzel and Andrea Ghez received the prize on Tuesday for their discoveries that have improved understanding of the universe, including work on black holes.
13. Chemistry
14. The Nobel Prize in Chemistry was jointly awarded on Wednesday to Emmanuelle Charpentier and Jennifer A. Doudna for their work on the development of Crispr-Cas9, a method for genome editing.
15. Literature
16. The Nobel Prize in Literature was awarded on Thursday to Louise Glück, one of America's most celebrated poets, ?for her unmistakable poetic voice that with austere beauty makes individual existence universal.?
17. Peace Prize
18. The Nobel Peace Prize was awarded on Friday to the World Food Program for its efforts to combat a surge in global hunger amid the coronavirus pandemic, which has swept around the world with devastating impact.
19. Economic Science
20. Paul R. Milgrom and Robert B. Wilson were awarded the Nobel in economic science on Monday for improvements to auction theory and inventions of new auction formats.

Detailed Review of Summarization Methods

Chan Young Park's slides:

<http://demo.clab.cs.cmu.edu/algo4nlp19/slides/summarization.pdf>

Reference: State of Art and Metrics

<http://nlpprogress.com/english/summarization.html>

Lecture 20: Concluding Comments

- We completed projects review – midway
 - Be mindful of project deliverable and dates
 - Take feedback into account
- Discussed automatic summarization
 - Extractive summarization is most common
 - Abstractive summarization is an ongoing research direction

Concluding Segment

Upcoming Classes

21	Mar 25 (Th)	Review: project presentations, Discussion	
22	Mar 30 (Tu)	Wellness Holiday	
23	Apr 1 (Th)	Text: Summary	
24	Apr 6 (Tu)	Text: Representation, Sentiment	
25	Apr 8 (Th)	Text: Sentiment, Visualization	
26	Apr 13 (Tu)	Case Study: Finance – Guest lecture	
27	Apr 15 (Th)	Paper presentations – Graduate students	Quiz 4 // Final assignment for Graduate students
28	Apr 20 (Tu)	Advanced: Bias and Trust Issues	
30	Apr 22 (Th)	Project presentations	
31	Apr 27 (Tu)	Reading day	



About Next Lecture – Lecture 23

Lecture 23: More on Text

- Word representation, language model
- Sentiment detection
- Very common, easy to implement and hard to get right!