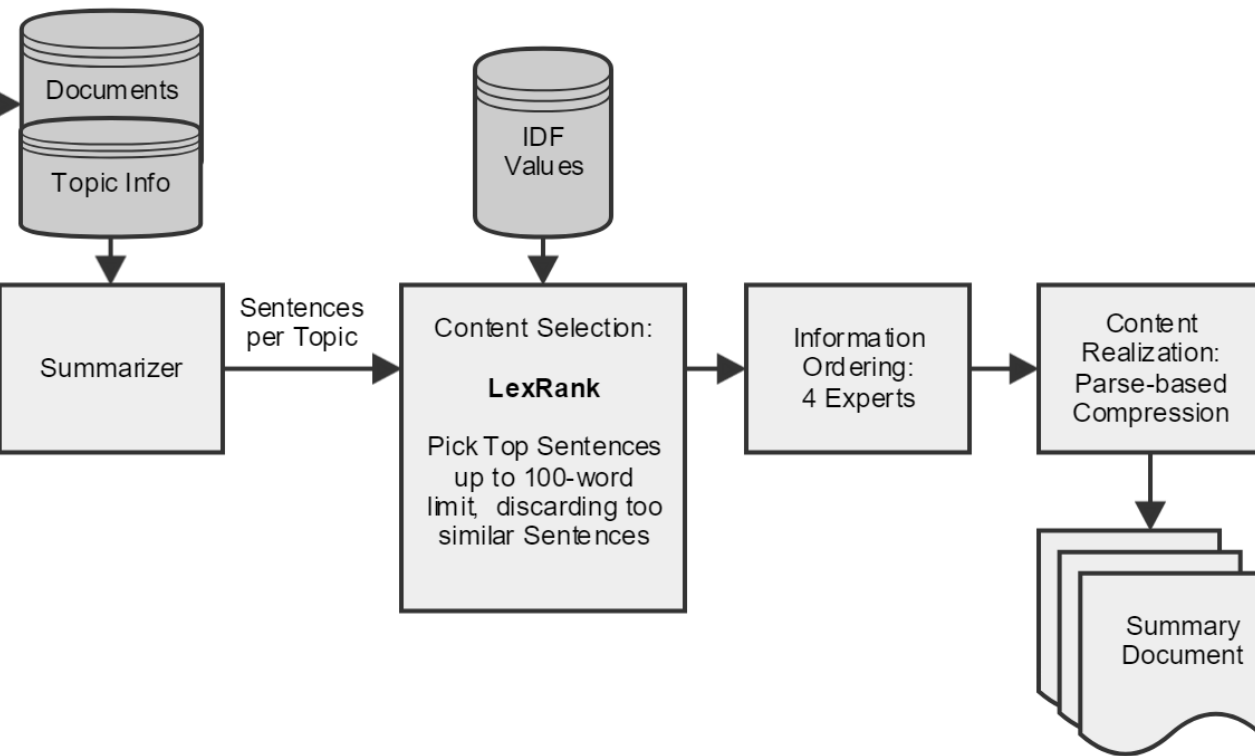


Multi-Document Summarization

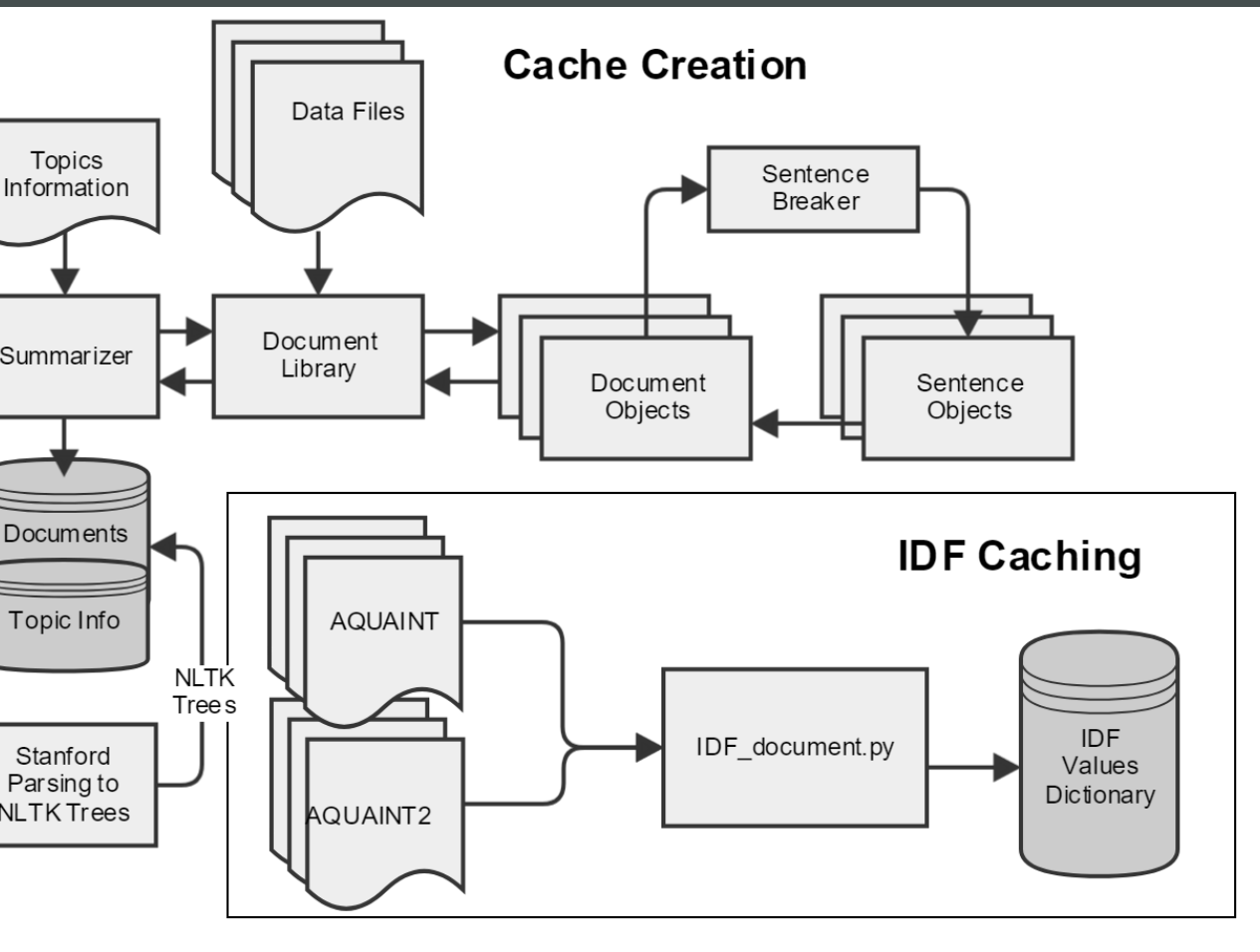
DELIVERABLE 4: CONTENT REALIZATION AND FINAL SYSTEM

TARA CLARK, KATHLEEN PREDDY, KRISTA WATKINS



System Architecture

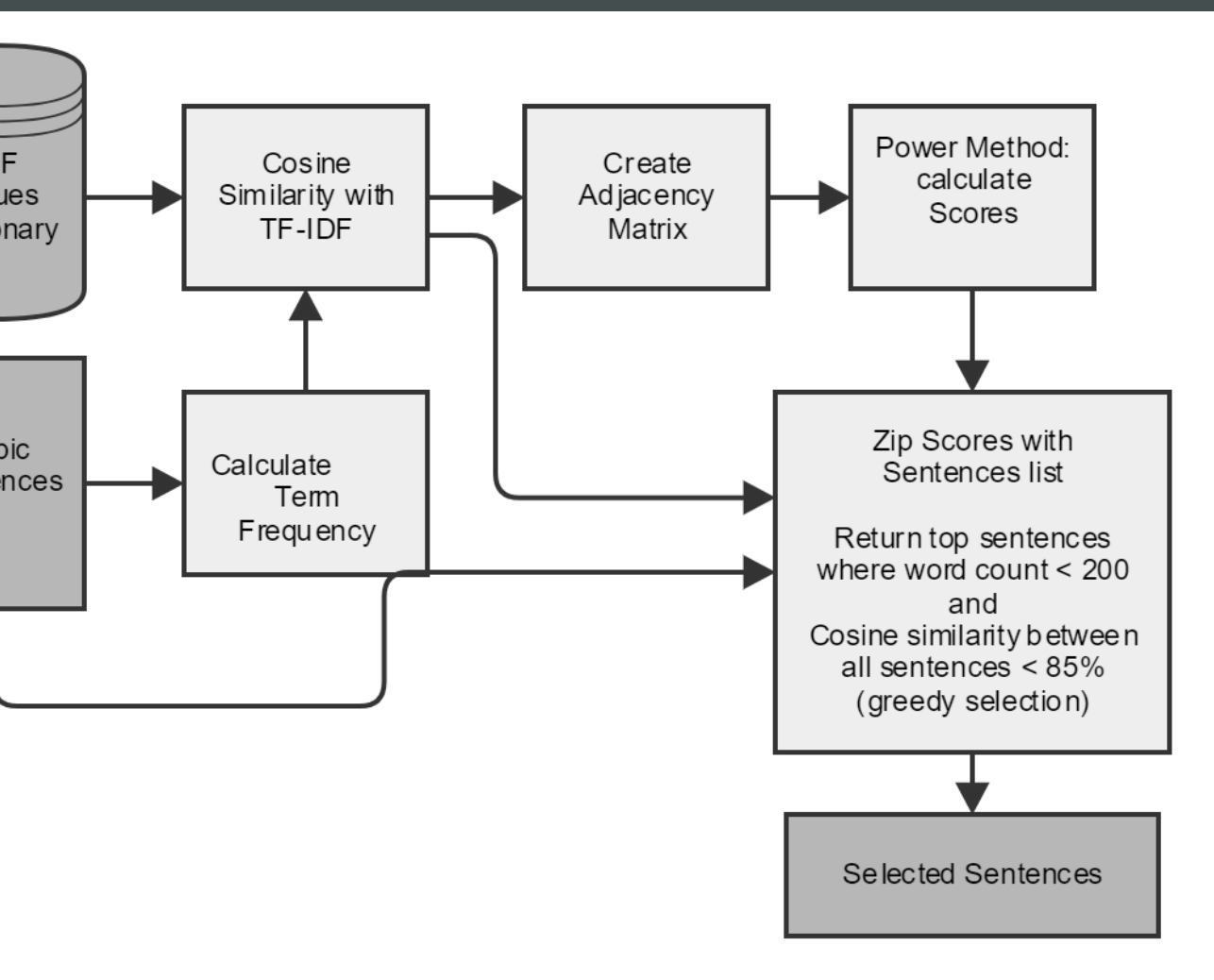
Our system is a collection of independent Python modules, linked together by the *Summarizer* module.



Cache Creation

Content Selection: Overview

- Input: Documents in a Topic
- Algorithm: Query-focused LexRank
- Output: List of best sentences, ordered by rank



Content Selection

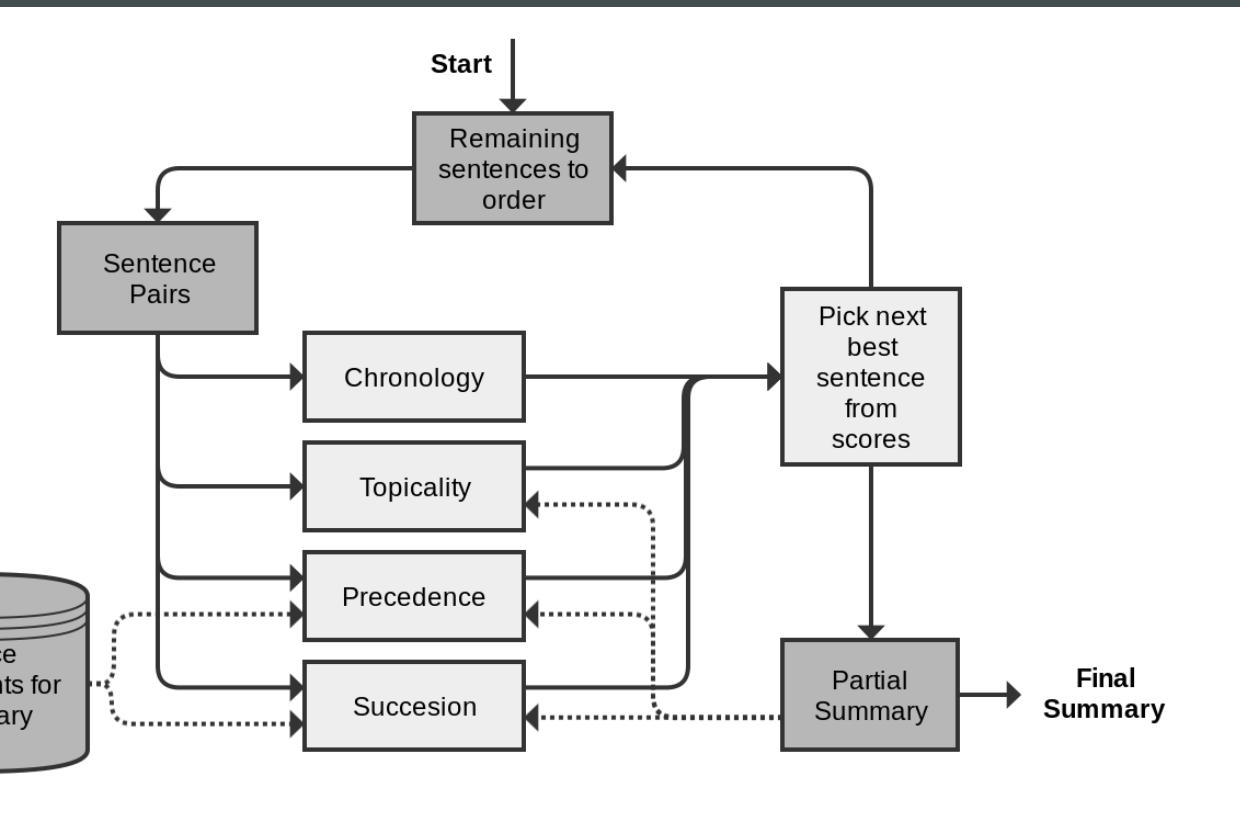
Architecture

D4 Additions

- Tried removing stopwords from the tf-idf calculation, but it made results worse
- Removed sentences less than 6 words long from consideration
 - Did not affect ROUGE
 - Seemed to clean out titles, etc., leading to better readability

Information Ordering

- Input: List of sentences from content selection
- Algorithm: Expert voting (Bollegata et al.)
- Output: List of ordered sentences

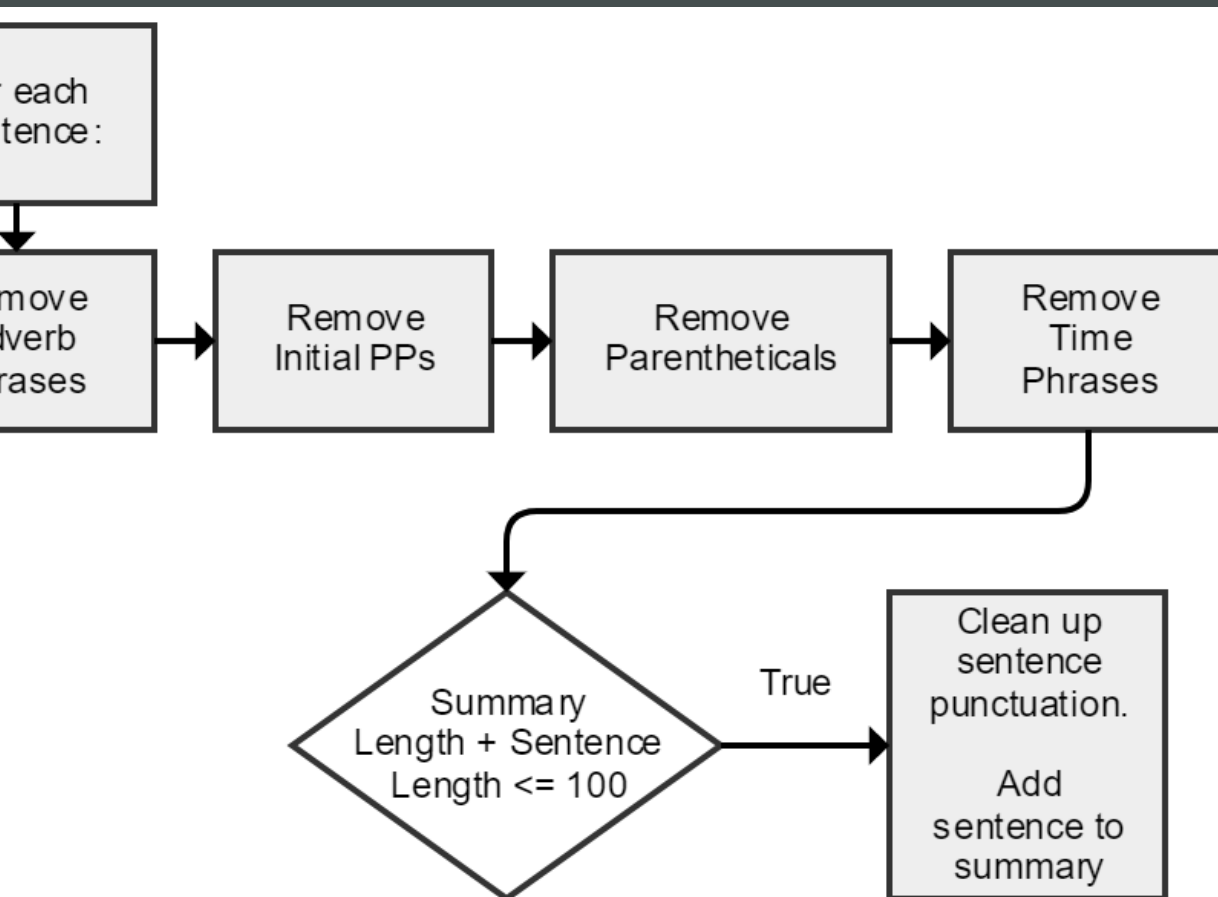


Information Ordering

Architecture

Content Realization

- Input: List of sentences from Information Ordering
- Trim the length of the summary to be 100 words, max
- Apply Compression Heuristics
- Output: Write each sentence on a new line to the output file



Content Realization

Architecture

Compression Heuristics

- Adverbial: Finds and removes any ADVP in the parse
- Initial prepositional: Finds and removes any sentence-initial prepositional phrases in the parse
- Parenthetical: Remove any text between parentheses
- Temporal: Finds and removes any prepositional phrases containing temporal vocabulary

Issues and Successes

- Scores went mysteriously down during the beginning of D4

Issues and Successes

- Content Realization heuristics
 - Adverbials:
 - ROUGE-1 + 0.00064
 - ROUGE-2 + 0.0035
 - ROUGE-3 - 0.00369
 - ROUGE-4 + 0.00419
 - Initial PPs:
 - ROUGE-1 + 0.00571
 - ROUGE-2 + 0.00598
 - ROUGE-3 + 0.00151
 - ROUGE-4 + 0.00515
 - Parentheticals: No change—no parentheses found in our summary sentences.
 - Temporal PPs:
 - ROUGE-1 - 0.00076
 - ROUGE-2 + 0.00338
 - ROUGE-3 + 0.00013
 - ROUGE-4 + 0.00449

Best Summary

Rouge 1: 0.45996

Rouge 2: 0.18431

- A major manufacturer of dog and cat food sold under Wal-Mart, Safeway, Kroger and other store brands recalled 60 million containers of wet pet food Friday after reports of kidney failure and deaths.
- An unknown number of cats and dogs suffered kidney failure and about 10 died after eating the affected pet food, Menu Foods said in announcing the North American recall.
- Product testing has not revealed a link explaining the reported cases of illness and death, the company said.
- Wheat gluten is a source of protein.

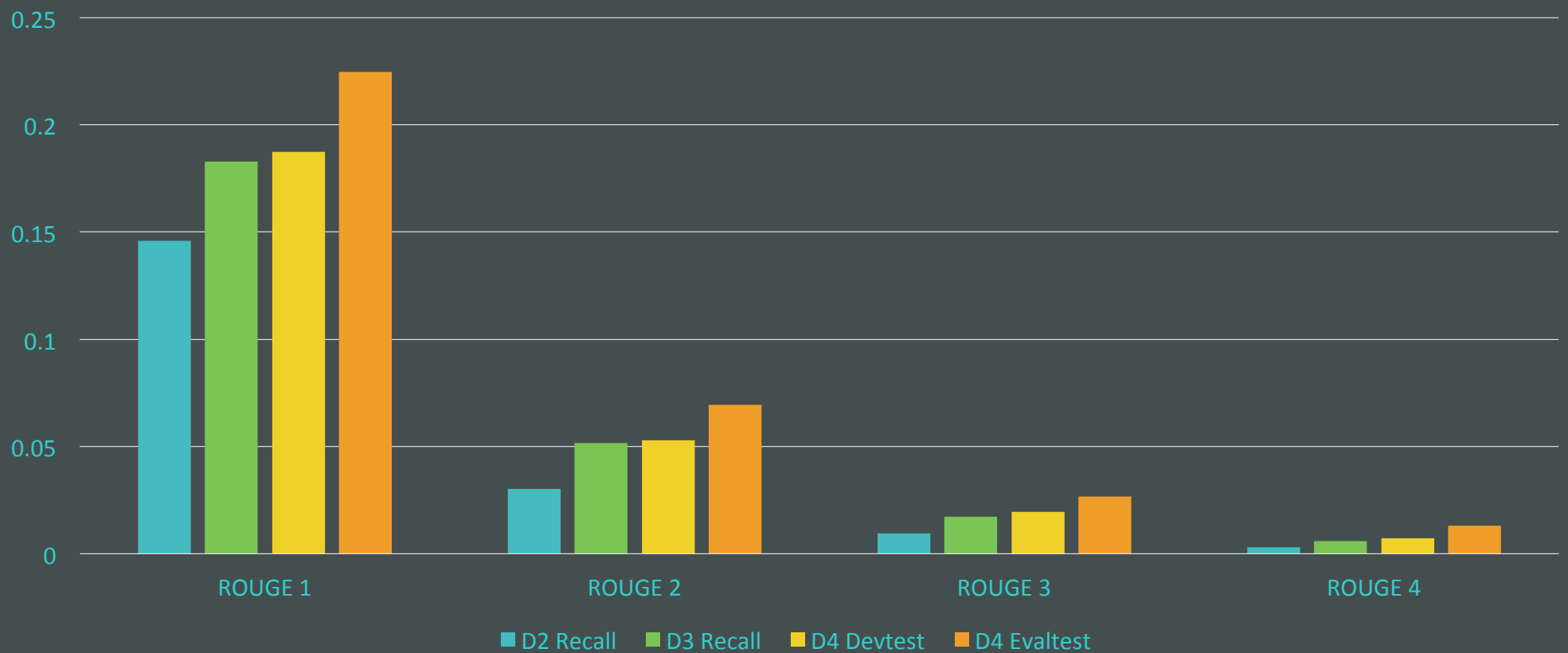
Worst Summary (Rouge 1: 0.04286)

- Saturday, Steven Toby settled down for a quick nap and woke up in a nightmare.
- The 51-year-old Shadow Hills man was headed, catching a little shut-eye his way to his job at Los Angeles City Hall.
- The clock read 6:02 a.m.
- A tremendous collision jarred him awake.
- His train had gone off the rails.
- ``You think, `Oh, I'll get through this and go on like, '''he said.
- The lights went out and heard a mass gasp.

Readable Poorly-Scoring Summary (4th worst Rouge 1: 0.10550)

- Is there no safe haven for President Bush?
- It happened through his news conference morning, between his 10th and 11th mentions of al-Qaida: A bird flew over president and deposited a wet, white dropping on upper left sleeve of his jacket.
- There was no evidence that Osama bin Laden was responsible for this particular attack, and -- who knows? -- maybe the terrorist leader believes the superstition that bird poop is good luck.
- Bush wiped the mess off with his bare hand.

Results



Results

	D2 Recall	D3 Recall	D4 Recall: Devtest	D4 Recall: Evaltest
ROUGE-1	0.14579	0.18275	0.18746	0.22452
ROUGE-2	0.03019	0.05149	0.05277	0.06956
ROUGE-3	0.00935	0.01728	0.0194	0.02658
ROUGE-4	0.00285	0.00591	0.00733	0.01304

Related Reading

Regina Barzilay, Noemie Elhadad, and Kathleen R. McKeown. 2002. Inferring strategies for sentence ordering in multidocument news summarization. *J. Artif. Int. Res.*, 17(1):35–55, August.

Danushka Bollegala, Naoaki Okazaki, and Mitsuru Ishizuka. 2012. A preference learning approach to sentence ordering for multi-document summarization. *Inf. Sci.*, 217:78–95, December.

Gunes Erkan and Dragomir R Radev. 2004. LexRank: Graph-based Lexical Centrality as Salience in Text Summarization. *Journal of Artificial Intelligence Research*, 22:457–479.

Ani Nenkova, Rebecca Passonneau, and Kathleen McKeown. 2007. The pyramid method: Incorporating human content selection variation in summarization evaluation. *ACM Trans. Speech Lang. Process.*, 4(2), May.

Jahna Otterbacher, Gunes Erkan, and Dragomir R. Radev. 2005a. Using random walks for question focused sentence retrieval. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, HLT '05*, pages 915–922, Stroudsburg, PA, USA. Association for Computational Linguistics.

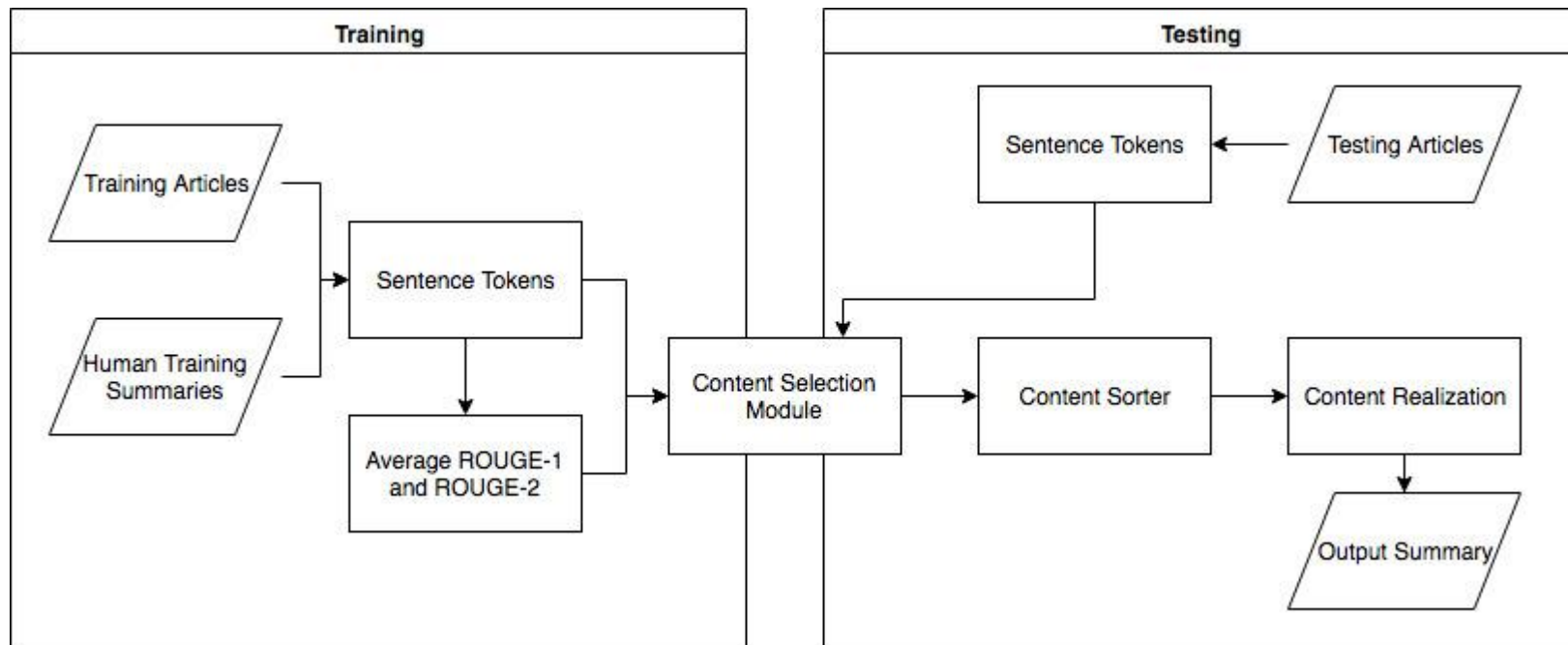
Karen Sparck Jones. 2007. Automatic summarising: The state of the art. *Inf. Process. Manage.*, 43(6):1449–1481, November.

Questions?

573 Project Report - D4

— Mackie Blackburn, Xi Chen, and —
Yuan Zhang

System Overview



Improvements in Content Selection

Larger background corpus for LLR

- Half of the New York Times corpus on Patas

Tweaking MLP regression

- 1 hidden layer of size 50

- Adaptive learning rate

Sentence Compression

Little to no effect on scores (R^2 -14%):

Ages

Dates/times

Attributions

Negative effect on scores (R^2 -26%):

Adjectives

Adverbs

Initial Conjunctions

Modifications in Content Realization

Sentence compression is introduced

In content realization, a modified greedy algorithm is applied:

- 1, while **compressed sentence length** does not exceed word limit:
- 2, pick the sentence with the highest score among candidates
- 3, unless the sentence's tf-idf similarity with candidates exceed threshold
($t < 0.4$)

Info Ordering: Review

As the result, we augment each summary sentence into a sentence group in the input documents by label spreading. Then we approximate sentence co-occurrence $CO_{m,n}$ by sentence group co-occurrence probability:

$$C_{m,n} = f(G_m, G_n)^2 / (f(G_m)f(G_n))$$

Here the $f(G_m, G_n)$ is the sentence group co-occurrence frequency within a word window and $f(G_m)$ is the sentence group co-occurrence frequency. This probability is about sentence groups' adjacency to each other.

Experiments on Info Ordering

Evaluation Dataset: 20 human extracted passages (of 3~4 sentences each) from training data, evaluate Kendall's tau on algorithm output vs human summaries.

Name	tau	Description
Adjacency 1	0.39	dim(word vector) = 50
Adjacency 2	0.41	dim(word vector) = 100
Adjacency 3	0.44	dim(para2vec) = 100
Adjacency 4	0.33	dim(word vector) = 200
Chronological	0.46	

Reflection on Info Ordering

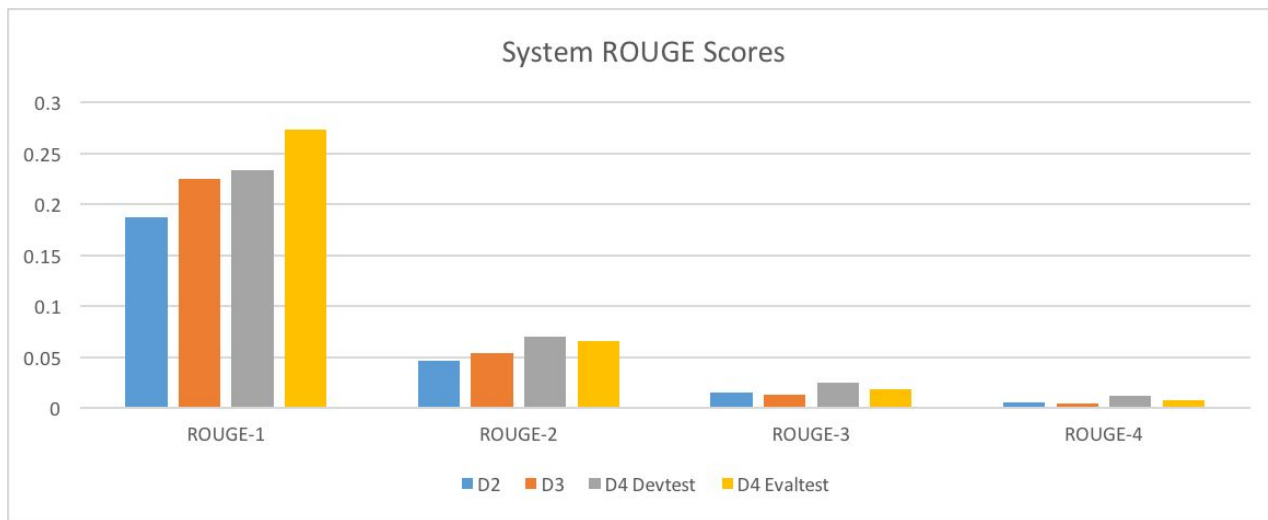
No word/sentence embedding based similarity can exceed chronological ordering. Still, they do better in different cases/test passages. Therefore, I assume a more efficient algorithm should take both into account. A more preferable solution is ordering/clustering by:

$$\text{Similarity}(s_1, s_2) =$$

$$t_1 * \text{semantics_diff}(s_1, s_2) + (1 - t_1) * \text{chronological_diff}(s_1, s_2)$$

This is just a variant of Bollegala et al(2012), 'A preference learning approach to sentence ordering for multi-document summarization' - combine different criterion together with machine learning based parameter adjusting.

Results



Model	ROUGE-1	ROUGE-2	ROUGE-3	ROUGE-4
D2	0.18687	0.04579	0.01503	0.00558
D3	0.22459	0.05354	0.01287	0.00424
D4 Devtest	0.23406	0.07048	0.02468	0.01198
D4 Evaltest	0.27357	0.06555	0.01845	0.00766

Issues and Successes

Minimal sentence compression has little to no effect on scores

Aggressive sentence compression has negative effects on scores and readability

Feature dependency

Ex. 1: “Parkinson’s disease”

D3

I saw it listed as the cause of death in an obituary. Exercise alone was enough to prevent the degeneration of brain cells in rats with **Parkinson's** disease, University of Pittsburgh researchers report. Today, a major treatment strategy is aimed at developing medicines to stop this abnormal protein from clumping. The research also identified a group of brain receptors in mice that appear to be responsible for nicotine addiction. It is a form of dementia that the National Institute of Neurological Disorders and Stroke calls dementia with Lewy bodies. The loss of cells that produce the neurotransmitter dopamine causes the telltale tremors, rigid and slow movements of **Parkinson's**. Distributed by the Los Angeles Times/Washington Post News Service

D4

Fox, who has **Parkinson's** disease, campaigned with Kerry in New Hampshire on Monday and filmed the ad after the event. The 84-year-old has **Parkinson's** disease, which makes it difficult for him to walk and to pronounce his words. Exercise alone was enough to prevent the degeneration of brain cells in rats with **Parkinson's** disease, University of Pittsburgh researchers report. The loss of cells that produce the neurotransmitter dopamine causes the telltale tremors, rigid and slow movements of **Parkinson's**. John Paul _ the most traveled pope in history _ cut back on his trips a few years ago. Investigators studied five families with a history of **Parkinson's** disease who lived in the Basque region of Spain and in England.

Ex. 2: “bird flu”

D3

While Hong Kong's imports mostly come from Shenzhen, Macao's are mainly from Zhuhai, Zhongshan and Jiangmen cities. A SAR government spokesman said the group, which comprises representatives from various government departments, will hold its first meeting Tuesday. A 54-year-old succumbed to the virus early this month, but a 2-year-old boy recovered after hospitalization in November. Although several cases were reported in Hong Kong, none was found in Macao so far. Macao residents are eating less birds these days although there has been no report of bird flu cases in Macao. The sheet of paper provided information of Influenza A H5N1, cautioning tourists to keep up a good body immunity.

D4

The "bird flu" has claimed four victims here, killing two, including a 54-year-old man who died Friday. A SAR government spokesman said the group, which comprises representatives from various government departments, will hold its first meeting Tuesday. Seven people have been infected so far, with two dead and two in critical conditions. The poultry imported from China's inland areas are from a different source from those imported by Hong Kong, they said. Local chicken farmers say that sales have dropped between 30 percent to 50 percent over the past weeks. Although several cases were reported in Hong Kong, none was found in Macao so far.

Ex. 3: “mad cow disease”

D3

The human form of **mad cow** disease is called variant Creutzfeldt-Jakob. The fatal brain-wasting disease is believed to come from eating beef products from cows struck with **mad cow** disease. **Tons of meat went to the market in violation of European norms", Valchovski, a virus expert and medical doctor, told AFP.** The money would come from the euro188 million (US\$235 million) set aside in 2005 to combat animal diseases in the EU. The vast bulk of them are elderly dairy cattle who would have eaten cattle-based feed in the 1980s. The likely vector of contamination for livestock was brain and nerve tissue mixed in animal feed.

D4

The human form of **mad cow** disease is called variant Creutzfeldt-Jakob. The fatal brain-wasting disease is believed to come from eating beef products from cows struck with **mad cow** disease. **Some 141 people are known to have died of vCJD in Britain.** Ireland banned the use of meat and bone meal as cattle feed, the suspected origin of **mad cow** disease, in 1990. It is estimated the **mad cow** crisis has cost the Canadian beef industry and rural economies about 5 billion US dollars. The money would come from the euro188 million (US\$235 million) set aside in 2005 to combat animal diseases in the EU. It icope with this difficult situation," she said.

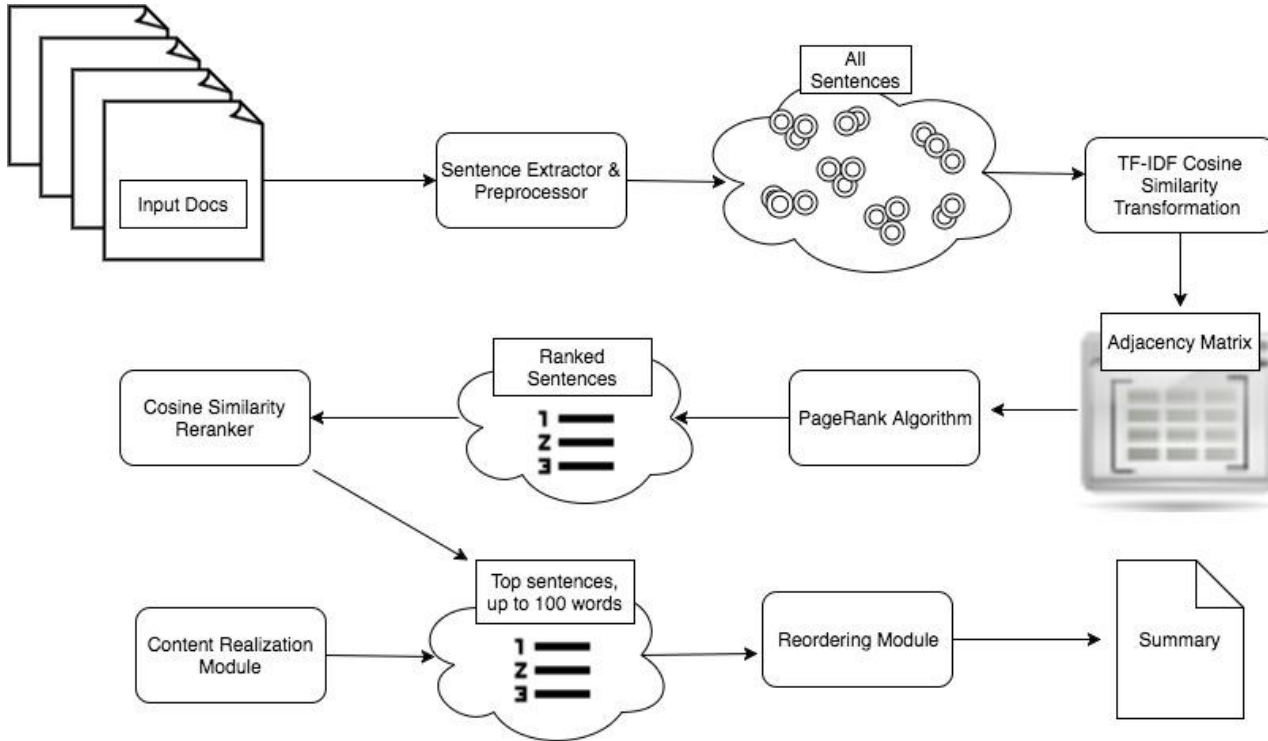
References

- [1] Ji Donghong and Nie Yu. Sentence ordering based on cluster adjacency in multi-document summarization. *The third international joint conference on natural language processing*, 2008.
- [2] G. Erkan and D. Radev. Lexrank: graph-based lexical centrality as salience in text summarization. *J. Artificial Intelligence Research*, 22(1):457-479, 2004.
- [3] J. Otterbacher G. Erkan and D. R. Radev. Using random walks for question-focused sentence retrieval. *Proceedings of Human Languages Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pp.915-922, 2005.
- [4] K. Hong and A. Nenkova. Improving the estimation of word importance for news multi- document summarization. in *Proceedings of EACL*, 2014.
- [5] P.E. Genest G. Lapalme and M. Yousfi-Monod. Hextac: The creation of a manual extractive run. *Proceedings of the Second Text Analysis Conference (TAC 2009)*, 2009.
- [6] A. Nenkova and K. McKeown. Automatic summarization. *Foundations and Trends in Information Retrieval*, 2011.

Multi-Document Summarization - D4

— Eslam Elsayy, Audrey Holmes, —
Masha Ivenskaya

System Architecture



Content Realization - What worked

- Removing sentence-initial adverbs and conjunctions
 - ~~“But~~ the new research, led by Baker, suggests that heat-tolerant algae may move in to replace strains lost in bleaching events.”
- Removing parenthetical expressions
 - Australian Prime Minister John Howard said Wednesday that his country would provide 1 billion Australian dollars ~~(about 764 million US dollars)~~ in loans and grants to assist Indonesia in its rebuilding after the Dec. 26 earthquake-tsunami disaster
- Removing ages
 - A judge heard motions Friday from lawyers for the two brothers, Deepak Kalpoe, ~~21~~, and Satish Kalpoe, ~~18~~, and a Dutch youth, Joran Van Der Sloot, ~~17~~

Content Realization - What didn't work

- Removing attribution phrases like “, he said”, “, according to”
 - this created several types of readability issues
- Removing noun appositives
 - problematic because of incorrect parses

Content Selection Improvements

Removing sentences with no named entities

Topic: Murder of JonBenet Ramsey

...	...	@@ -1,5 +1,4 @@
1	1	Hunter took the JonBenet case to the grand jury shortly after a former Boulder police detective on the case and three former friends of the Ramseys publicly demanded that Colorado's governor, Roy Romer, replace Hunter on the case with a special prosecutor.
2	2	Have been no arrests, although police have said JonBenet's parents, John and Patsy Ramsey, are under suspicion.
3	3	To push the Ramsey case along, prosecutors from neighboring counties have lent to Hunter's office prosecutors with experience in homicides and grand juries.
4		- A new grand jury was impaneled here two months ago.
5		- It is amazing the lack of urgency.''
	4	+ Boulder Police Chief Mark Beckner has said they are under an umbrella of suspicion.''

Rouge1: 0.314  0.349 (+0.035)

Content Selection Improvements - I

Removing sentences with no named entities

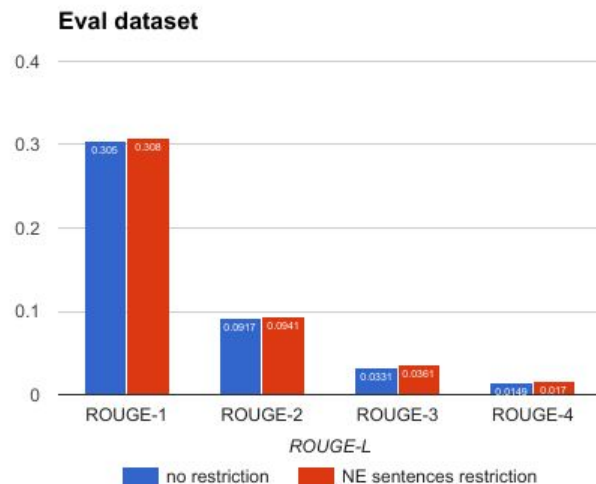
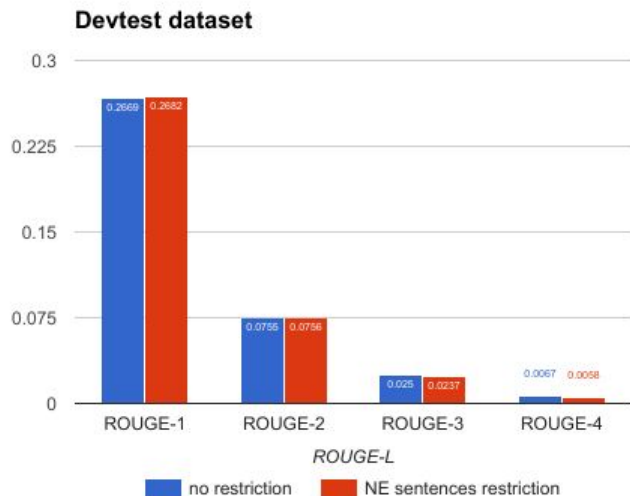
Topic: Bernard Madoff scandal

...	...	@@ -2,4 +2,4 @@ Santander may face \$3.1 billion in losses through its Optimal Investment Service
2	2	It is the second private Swiss bank to acknowledge having investments in funds managed by Madoff.
3	3	The bank said it had also invested 17 million euros of its own funds in Madoff products.
4	4	On Thursday, Madoff was charged with securities fraud after confessing to his sons that his business was a Ponzi scheme, according to a complaint filed by the Securities and Exchange Commission.
5		- He estimated that his investors had lost \$50 billion.
	5	+ Madoff eventually attracted billions of dollars from investors.

Rouge1: 0.331  0.323 (-0.008)

Content Selection Improvements - Analysis

Removing sentences with no named entities



Information Ordering Experiments

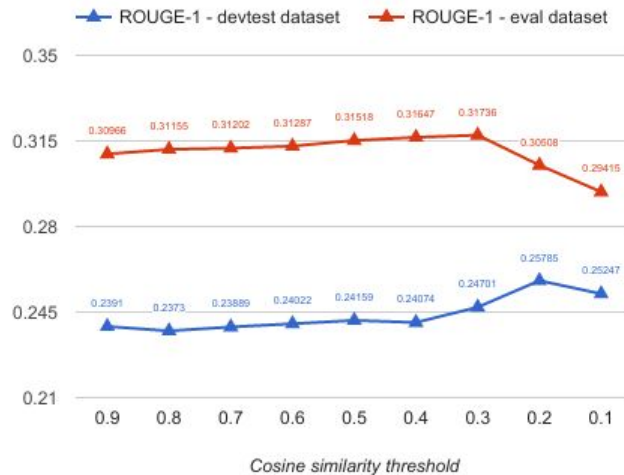
Effect of cosine similarity threshold on the rouge scores

Devtest dataset:

- Best result at threshold = 0.2

Evaluation dataset:

- Best result at threshold = 0.3



Information Ordering Experiments - I

Topic: Cyclone Sidr hits Bangladesh

Id	Sentence
1	The area is close to Bangladesh's border with the Indian state of West Bengal, which is also expecting highly destructive winds and torrential rains.
2	A powerful cyclone with strong winds started pounding on Bangladesh's south and southwestern coast from Thursday evening.
3	Officials in both Bangladesh and India have been racing to evacuate hundreds of thousands of people from the area over the past 48 hours.
4	Cyclone Sidr, described as the worst storm in years to hit low-lying and disaster-prone Bangladesh, crashed into the southwestern coast Thursday night before sweeping north over the capital Dhaka.
5	Six water treatment systems were mobilized.

Cosine Sim ordering: 1 > 2 > 3 > 4 > 5

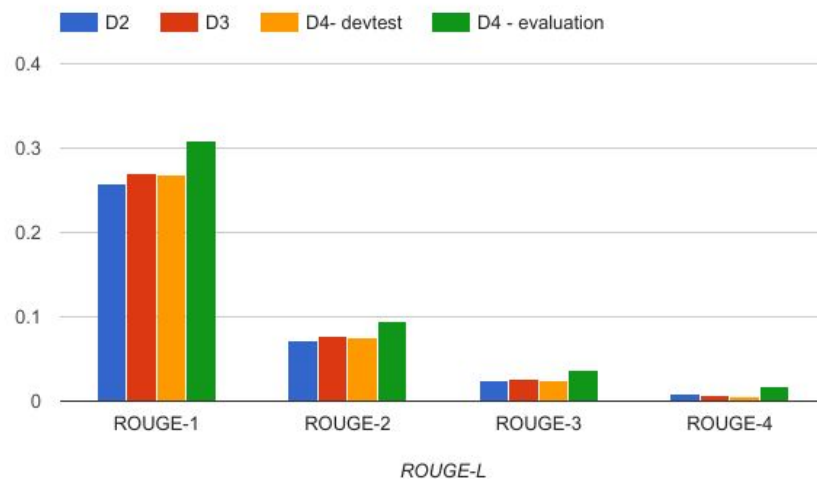
Cosine + entity grid: 4 > 2 > 1 > 3 > 5

Cosine + Cohesion expert: 1 > 3 > 4 > 2 > 5

Cosine + Chronological expert: 1 > 2 > 3 > 4 > 5

Rouge Scores

ROUGE-L	D2	D3	D4-dev	D4-eval
ROUGE-1	0.25785	0.27056	0.26828	0.308
ROUGE-2	0.07108	0.07684	0.07581	0.09428
ROUGE-3	0.02438	0.02596	0.02396	0.03601
ROUGE-4	0.00847	0.00739	0.00591	0.01696



References

- [1] Radev, Dragomir R., et al. "MEAD-A Platform for Multidocument Multilingual Text Summarization." *LREC*. 2004.
- [2] Erkan, Günes, and Dragomir R. Radev. "Lexrank: Graph-based lexical centrality as salience in text summarization." *Journal of Artificial Intelligence Research* 22 (2004): 457-479.
- [3] Lin, Chin-Yew. "Rouge: A package for automatic evaluation of summaries." *Text summarization branches out: Proceedings of the ACL-04 workshop*. Vol. 8. 2004.
- [4] Barzilay, Regina, and Mirella Lapata. "Modeling local coherence: An entity-based approach." *Computational Linguistics* 34.1 (2008): 1-34.
- [5] Jenny Rose Finkel, Trond Grenager, and Christopher Manning. 2005. Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling. *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL 2005)*, pp. 363-370
- [6] Dan Klein and Christopher D. Manning. 2003. [Accurate Unlexicalized Parsing](#). *Proceedings of the 41st Meeting of the Association for Computational Linguistics*, pp. 423-430.

Questions ?

Automatic Summarization System

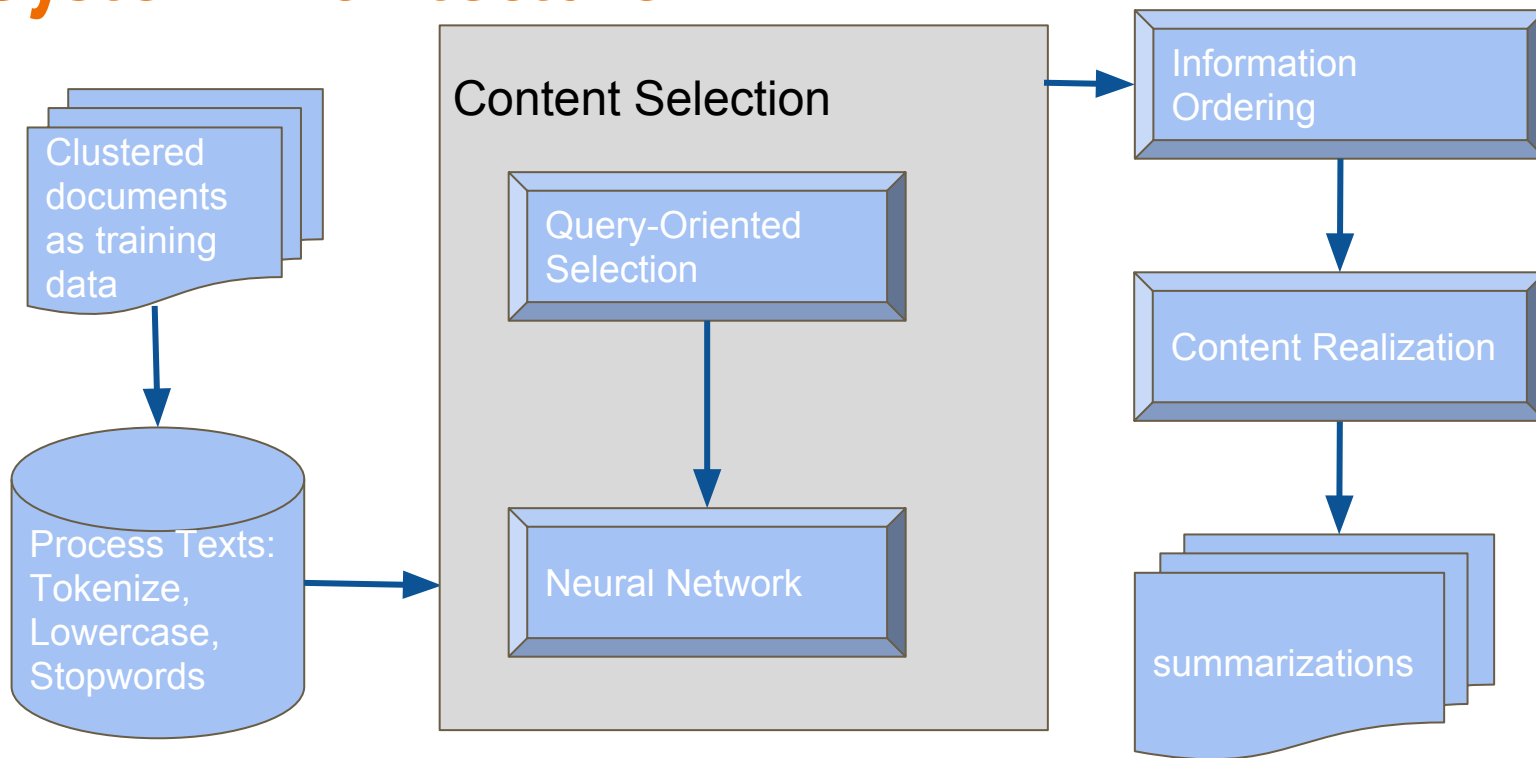
DELIVERABLE 4

— Wenxi Lu, Yi Zhu, Meijing Tian —

Outline

- System Architecture
- Content Selection
- Information Ordering
- Sentence Compression
- Results
- Issues and Discussion

System Architecture



Content Selection

- Query-oriented lexrank score
- Algorithm: Otterbacher 2005
- Output: Chronologically ordered sentences with high lexrank score
- Threshold: 30, 50, 100, 150 (sentences per topic)

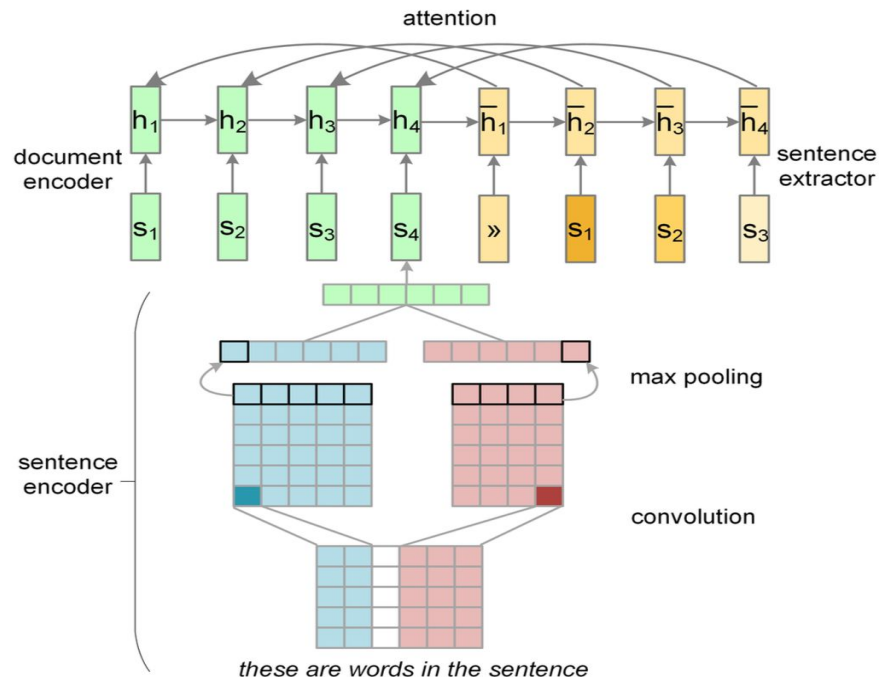
Content Selection

- Neural Network (Changes)

- Input
 - Chronologically ordered text
- Output all sentences with label 1
 - Precision > Recall
 - R3 and R4 high (Coherence)
- Different Hyperparameters
 - Max Sentence Length: 50
 - Max Sentence Number: 100

$$\bar{\mathbf{h}}_t = \text{LSTM}(p_{t-1}\mathbf{s}_{t-1}, \bar{\mathbf{h}}_{t-1})$$

$$p(y_L(t) = 1|D) = \sigma(\text{MLP}(\bar{\mathbf{h}}_t : \mathbf{h}_t))$$

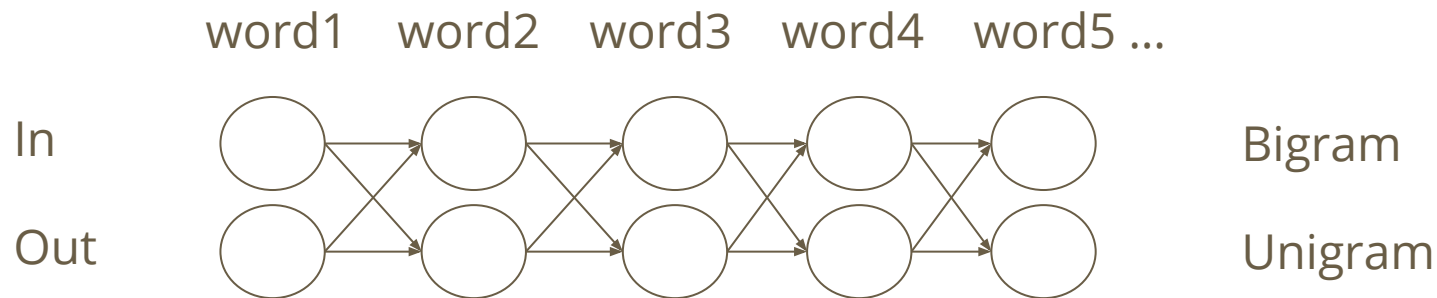


Information Ordering

- Algorithm: Barzilay et al, 2002
- Majority Ordering
- Chronological Ordering

HMM Sentence Compression

HMM Hedge (Zajic, David, et al., 2007)



HMM Sentence Compression

Parameter Tuning

- “Clumpiness” length & weight (contiguous block of words)
- Verb tag set weight ["VB", "VBD", "VBG", "VBN", "VBP", "VBZ"]
- Noun tag set weight ["NN", "NNP", "NNPS", "NNS", "PRP", "PRP\$"]
- Adverb tag set weight ["RB", "RBR", "RBS"]
- Preposition tag set weight ["IN", "TO"]

clumpPara	clumpLen	vbPref	nnPref	advPref	prepPref
1	4	0.3	0.3	-1.5	-0.3

HMM Sentence Compression

Post compression clean-up

- Remove 2nd IN in double INs
- Remove IN in IN-CC sequence
- Remove DT when next POS is not noun tag set, adjective tag set or number tag set
- If only INs and/or DTs between 2 CCs, only keep the 1st CC

*IN: preposition or subordinating conjunction, CC: coordinating conjunction

HMM Sentence Compression

Original

he said that the company had received an **undisclosed** number of owner complaints that **dogs** and cats were vomiting and **suffering** kidney failure after eating its products

Compressed

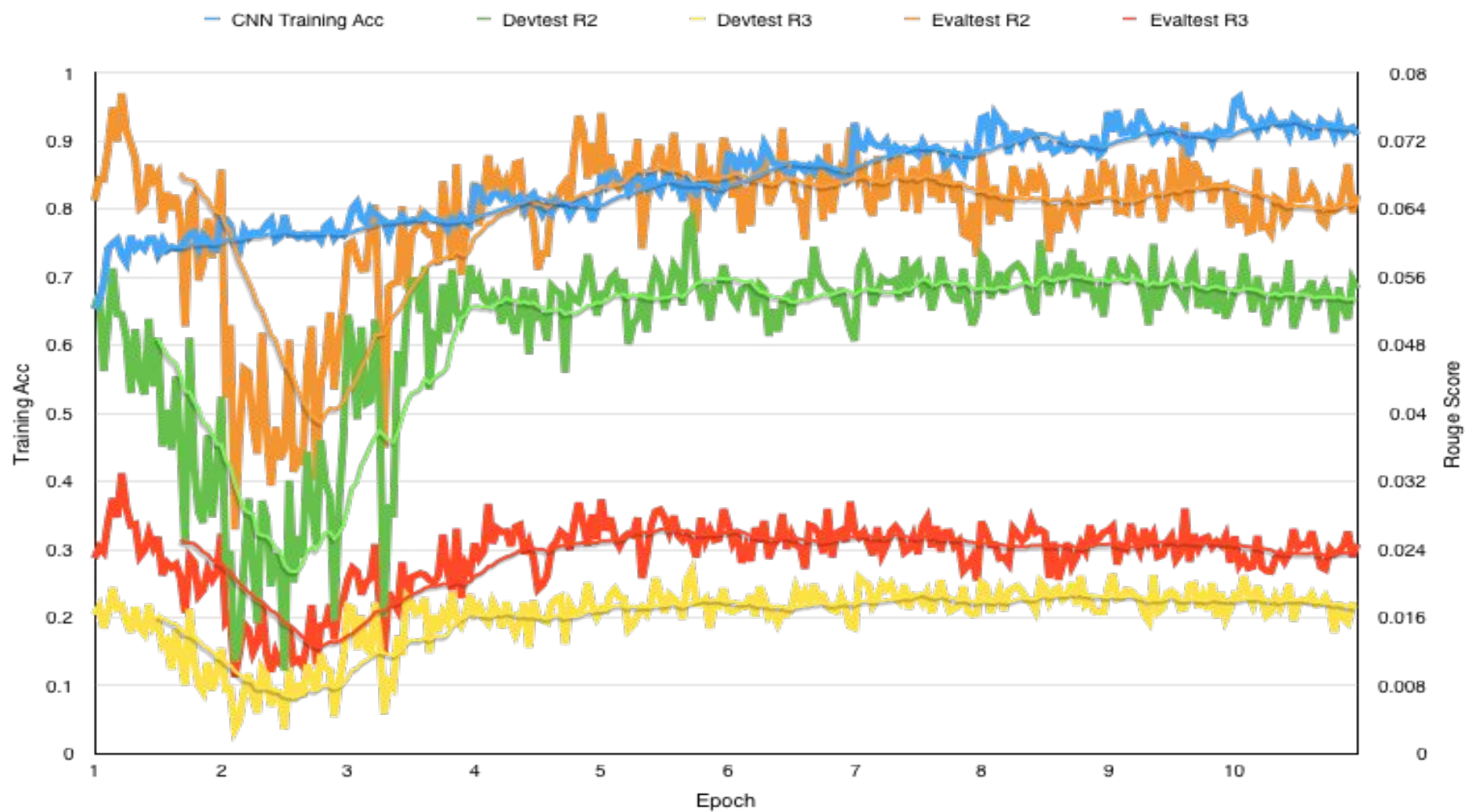
he said that company had received an number of owner complaints that and cats were vomiting and kidney failure after eating its products

Rule Sentence Compression

- Minor pruning of the sentences
- No major effect on the readability
 - Adverbials
 - Bracket Content : (U.S. 83 million)
 - Trailing Attribution : , the officer said

Results

- Training Data
 - CNN & Dailymail single document summarization
- Devtest
 - CNN
 - 30 sentences per document set
- Evaltest
 - Dailymail
 - 50 sentences per document set



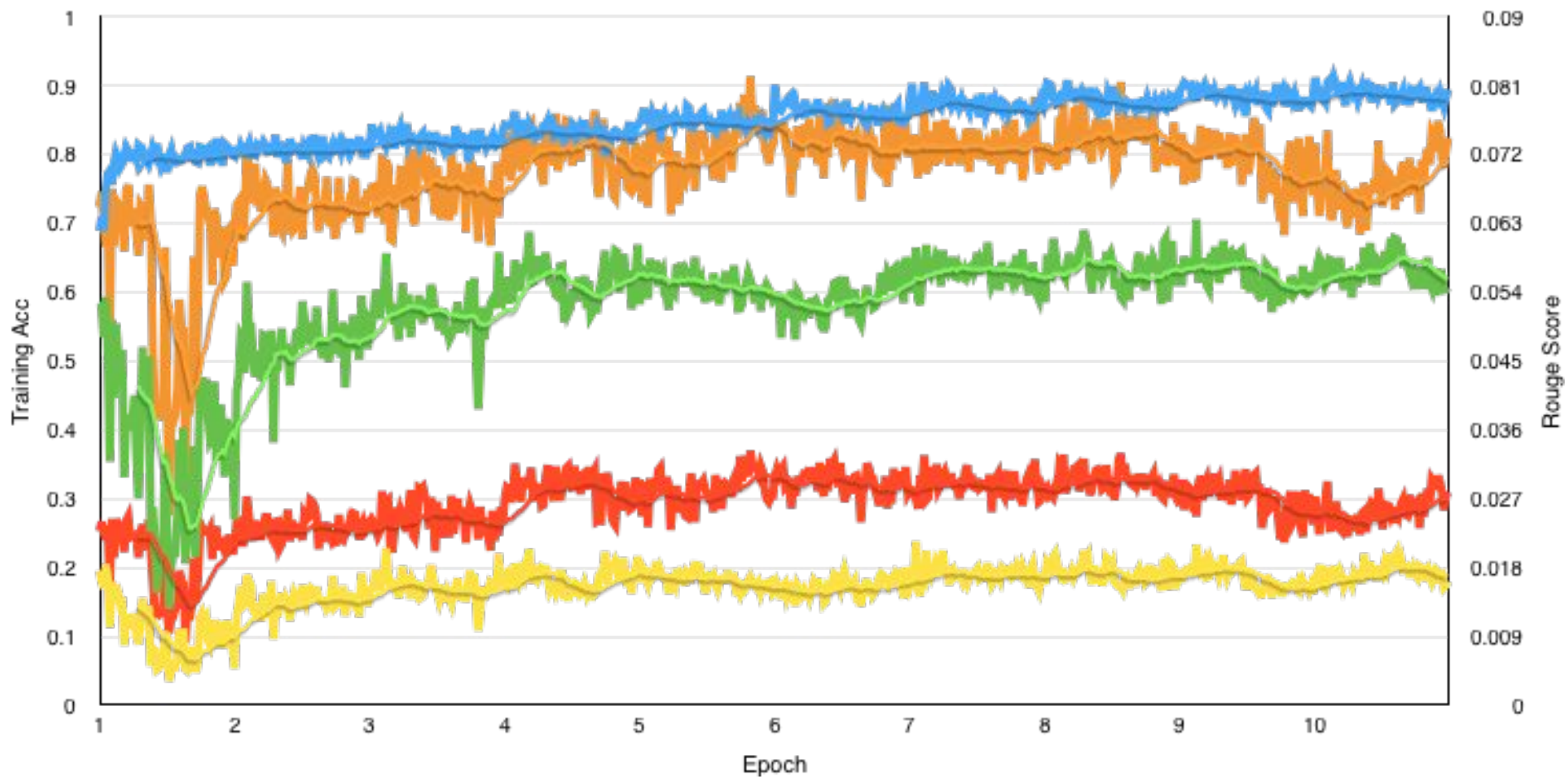
DailyMail Training Acc

Devtest R2

Devtest R3

Evaltest R2

Evaltest R3



	R1	R2	R3	R4
D2 (LR)	0.1935	0.0501	0.0167	0.0057
D2 (NN)	0.2287	0.0565	0.0154	0.0039
D3 (NN)	0.2079	0.0603	0.02079	0.0084
D3 (NN + LEX + MO)	0.1743	0.0387	0.01178	0.0041
D4 DEV (LEX)	0.19495	0.05308	0.01845	0.00484
D4 DEV (LEX + NN)	0.22572	0.06293	0.02152	0.00718
D4 DEV (LEX + NN + CO)	0.22640	0.06065	0.01827	0.00540
D4 DEV (LEX + NN + HMM)	0.23729	0.05508	0.01539	0.00390
D4 DEV (LEX + NN + R)	0.23178	0.06518	0.02205	0.00757
D4 EVAL (LEX)	0.24779	0.06216	0.02141	0.01056
D4 EVAL (LEX + NN)	0.26358	0.08225	0.03328	0.01710
D4 EVAL (LEX + NN + CO)	0.28071	0.08776	0.03327	0.01625
D4 EVAL (LEX + NN + CO + HMM)	0.29522	0.08063	0.02591	0.01033
D4 EVAL (LEX + NN + CO + RULE)	0.28836	0.09110	0.03556	0.01733

Examples

NN + CO

a gunman shot girls in the head `` execution style " at an amish school in pennsylvania state on monday , killing four and **wounding** at least six others , police and officials said .
miller confirmed three dead at the scene .

NN + CO + HMM

a gunman shot girls the head execution style at amish school in pennsylvania state monday killing four and at least six others police officials said
miller confirmed three dead the scene

NN + CO + RULE

a gunman shot girls in the head `` execution style " at an amish school in pennsylvania state on monday , killing four and wounding at least six others
miller confirmed three dead at the scene .

Issues & Discussion

- NN is not working very well
 - Discrepancy
- HMM fluency (lowered R2, R3 and R4 scores)
- Rule-based sentence compression works well
 - Non-fully abstractive NN will also work well?

References

- Regina Barzilay, Noemie Elhadad, and Kathleen McKeown. 2002. Inferring strategies for sentence ordering in multi-document news summarization. *Journal of Artificial Intelligence Research* 17:35–55.
- Gunes Erkan and Dragomir R Radev. 2004. Lexrank: Graph-based lexical centrality as salience in text summarization. *Journal of Artificial Intelligence Research* 22:457–479.
- Jahna Otterbacher, Gunes Erkan, and Dragomir Radev. 2005. Using random walks for question-focused sentence retrieval. *Journal of Artificial Intelligence Research* .
- Samy Bengio, Oriol Vinyals, Navdeep Jaitly, and Noam Shazeer. 2015. Scheduled sampling for sequence prediction with recurrent neural networks. In *Advances in Neural Information Processing Systems 28*, pages 1171–1179. Curran Associates, Inc.
- Zajic, David, et al. "Multi-candidate reduction: Sentence compression as a tool for document summarization tasks." *Information Processing & Management* 43.6 (2007): 1549-1570.