ECE594: Mathematical Models of Language

Spring 2022

Lecture 8: Text Summarization

Recap: Text Summarization

- Extractive summarization
  - summary is a subset of original text
- Abstractive summarization
  - summary is paraphrase of original text

Fourscore and seven years ago our fathers brought forth on this continent a new nation, conceived in liberty, and dedicated to the proposition that all men are created equal. Now we are engaged in a great civil war, testing whether that nation, or any nation so conceived and so dedicated, can long endure. We are met on a great battle- field of that war. We have come to dedicate a portion of that field as a final resting-place for those who here gave their lives that this nation might live. It is altogether fitting and proper that we should do this. But, in a larger sense, we cannot dedicate...we cannot consecrate...we cannot hallow... this ground. The brave men, living and dead, who struggled here, have consecrated it far above our poor power to add or detract. The world will little note nor long remember what we say here, but it can never forget what they did here. It is for us, the living, rather, to be dedicated here to the unfinished work which they who fought here have thus far so nobly advanced. It is rather for us to be here dedicated to the great task remaining before us...that from these honored dead we take increased devotion to that cause for which they gave the last full measure of devotion; that we here highly resolve that these dead shall not have died in vain; that this nation, under God, shall have a new birth of freedom; and that government of the people, by the people, for the people, shall not perish from the earth.

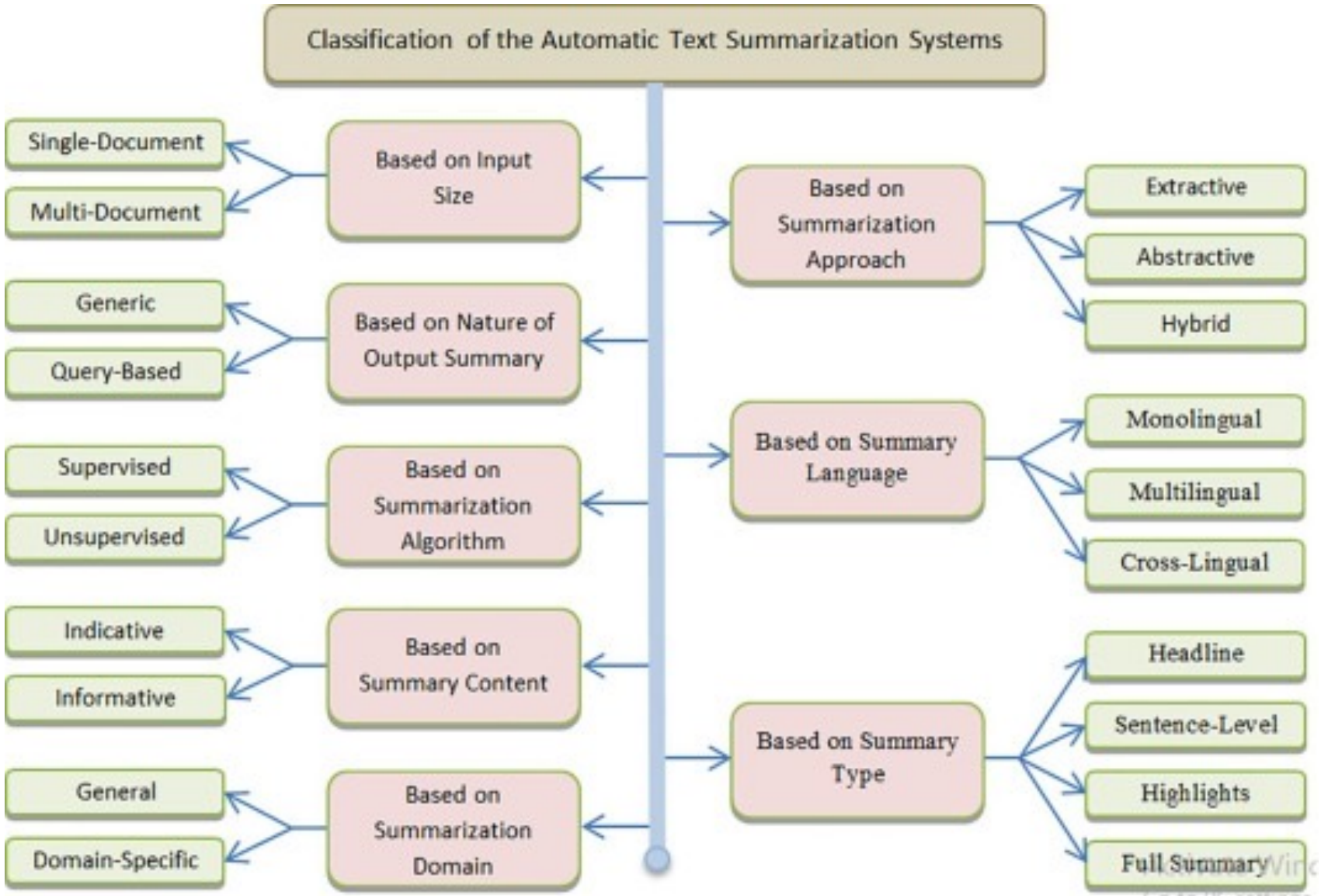
Figure 23.12 The Gettysburg Address. Abraham Lincoln, 1863.

Extract from the Gettysburg Address:
Four score and seven years ago our fathers brought forth upon this continent a new nation, conceived in liberty, and dedicated to the proposition that all men are created equal. Now we are engaged in a great civil war, testing whether that nation can long endure. We are met on a great battlefield of that war. We have come to dedicate a portion of that field. But the brave men, living and dead, who struggled here, have consecrated it far above our poor power to add or detract. From these honored dead we take increased devotion to that cause for which they gave the last full measure of devotion — that government of the people, by the people for the people shall not perish from the earth.
Abstract of the Gettysburg Address:
This speech by Abraham Lincoln commemorates soldiers who laid down their lives in the Battle of Gettysburg. It reminds the troops that it is the future of freedom in America that they are fighting for.

Figure 23.13 An extract versus an abstract from the Gettysburg Address (abstract from Mani (2001)).


EDITION: U.S. | INTERNATIONAL | MÉXICO
Set edition preference
Home Video NewsPulse U.S.
Egypt's military dis
suspends constitut
By the CNN Wire Staff
February 13, 2011 2:44 p.m. EST
Egypt suspends constitution
STORY HIGHLIGHTS
NEW: Banks are shuttered until Wednesday as protests force top banker's resignation
NEW: ElBaradei urges generals to "come out of their headquarters"
NEW: Stock exchange to freeze transactions from officials being investigated
Egypt's ambassador says the military will run a "technocratic" government until elections
Sameh Shoukry, Egypt's ambassador to the United States, said Sunday that the generals have made restoring security and reviving the economy its top priorities.
"This current composition is basically a technocratic government to
HOME PAGE TODAY'S PAPER VIDEO MOST POPULAR TIMES TOPICS
The New York Times
WORLD U.S. N.Y. / REGION BUSINESS TECHNOLOGY
AFRICA AMERICAS ASIA PACIFIC EUROPE MIDDLE EAST
Only at BlackBerry
Get the ScoreMobile Super App
Egyptian Military Dissolves Parlian
By ANTHONY SHADID
Published: February 13, 2011
CAIRO — The Egyptian military consolidated its control over what it has called a democratic transition from nearly three decades of President Hosni Mubarak's authoritarian rule, dissolving the feeble Parliament, suspending the Constitution and calling for elections in six months in sweeping steps that echoed protesters' demands.
Multimedia
TAHRIR SQUARE
The statement by the Supreme Council of the Armed Forces, read on television, effectively put Egypt under direct military authority, thrusting the country into territory uncharted since republican Egypt was founded in 1952. Though enjoying popular support, the military must now
TWITTER
SIGN IN TO E-MAIL
PRINT
SINGLE PAGE
REPRINTS
SHARE
NOW PLAYING EVERYWHERE
THE WALL STREET JOURNAL. MIDDLE EAST
Mideast Unrest Spreads
Protests Target Iran, Bahrain, Libya; Egypt
Article Video Slideshow
CAIRO—As Egypt's new military leadership suspended the constitution, dissolved parliament and promised fresh elections, demands for similar political reform swept across the Arab world—from Libya to Iran—following the resignation of President Hosni Mubarak.
USA TODAY | World
Ideology shadows fight to rule t
CAIRO — Ali Abdel Fattah, spokesman for the long-banned
… 27,000+ more

Text Summarization



- Generic summarization:
  - Summarize the content of a document
- Query-focused summarization:
  - summarize a document with respect to an information need expressed in a user query.
  - a kind of complex question answering:
    - Answer a question by summarizing a document that has the information to construct the answer

Extractive Summarization

- Select units from the original
  - Typically sentences
  - No simplification/rewriting

- Baseline
  - Extract the first few sentences (news genre)

Extractive Summarization

- Long history
  - Baxendale (1958)
  - Luhn (1958; technical documents)

- Heuristics
  - Position of sentences
    - Analyzed 200 paragraphs; first and last are topic sentences
  - Sentences with content terms (frequency/uniqueness)
  - Cue words (*hardly, significant, impossible*)

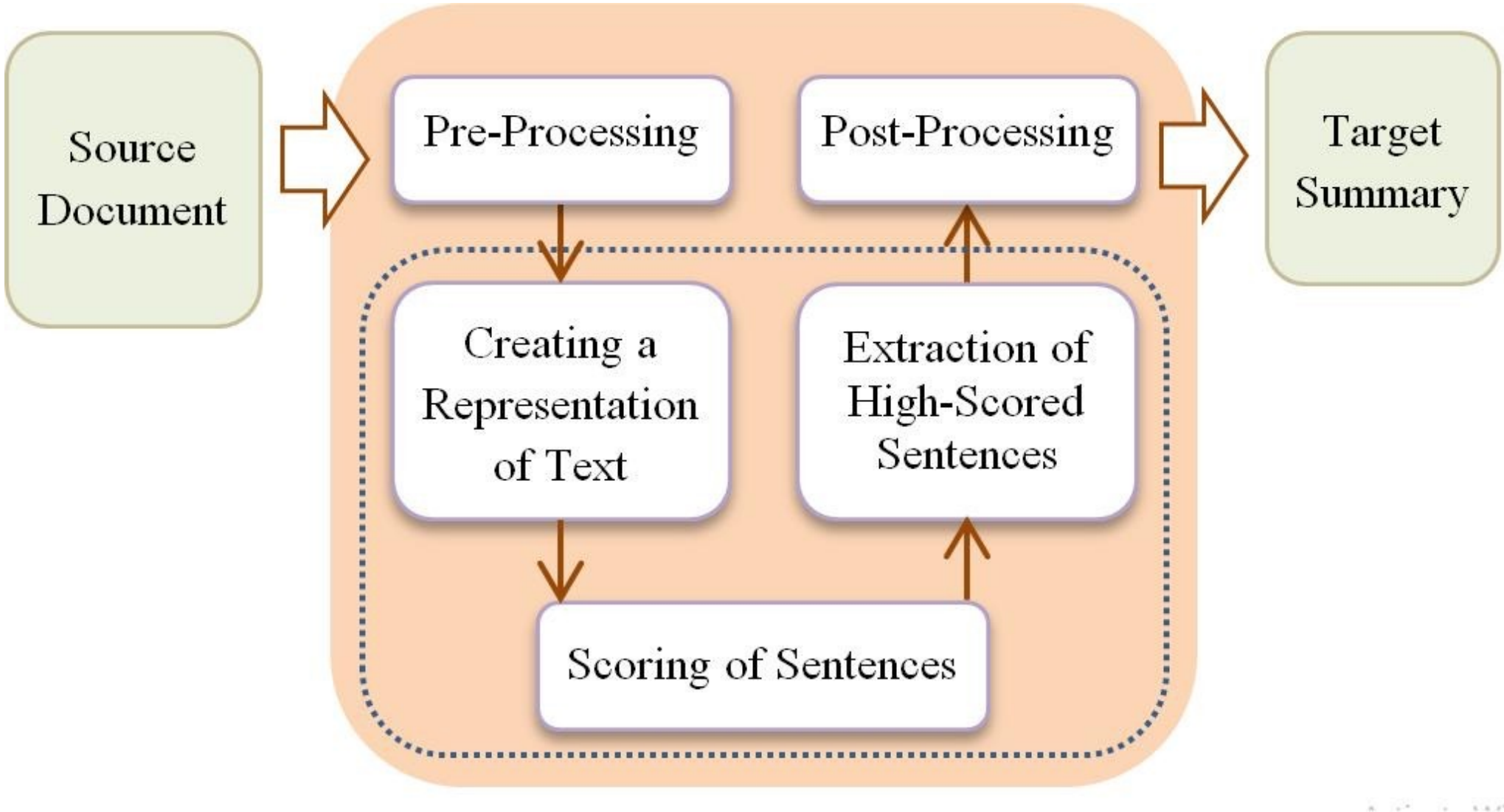
Extractive Summarization

- Problems
  - Paice (1990)
    - Lack of balance (e.g., single views)
    - Lack of cohesion (antecedent not mentioned/incorrectly cited)
- Solutions
  - Rhetorical structure theory
  - Anaphors
    - *That:* nonanaphoric if preceded by a research verb (demonstrated)
      : nonanaphoric if followed by pronoun, article, quantifier

Summarization Tasks

- Content selection
  - Choose sentences to extract
- Information ordering
  - Order sentences
- Realization
  - Cleanup and present

Text Summarization



Intermediate Representation

- Topic representation approaches convert the text to an intermediate representation interpreted as the topic(s) discussed in the text

Content Selection Methods (Topic signature)

- Topic signature
  - Set of salient terms
  - Computed using tf-idf
  - Rank words by tf-idf

Content Selection Methods (Topic signature)

- Tf-idf
  - Frequently occurring terms reflect meaning of document > less frequent terms
  - Terms limited to few documents discriminate those documents from the rest

$$w_{i,j} = \mathrm{tf}_{i,j} \times \mathrm{idf}_i \qquad \mathrm{idf}_i = \log\left(\frac{N}{n_i}\right)$$

Content Selection Methods (Graph-Based)

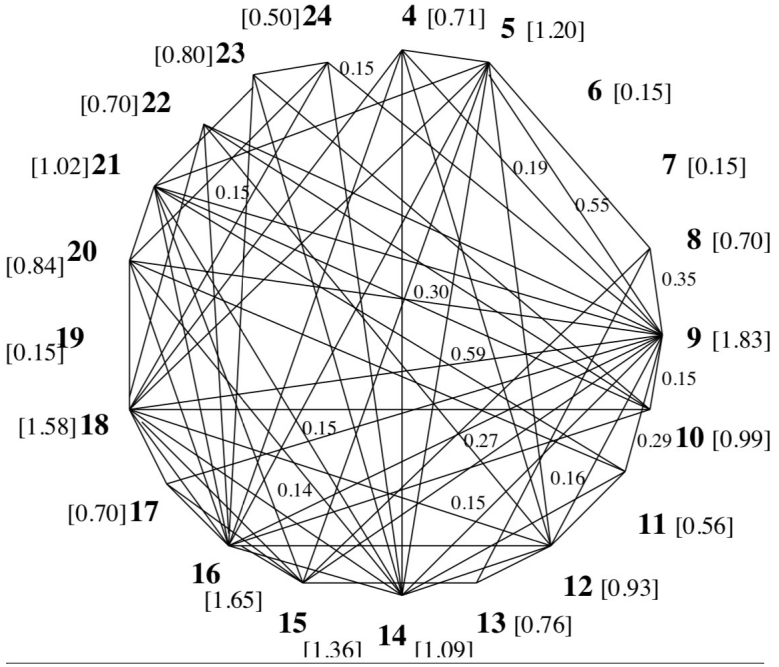
- Sentences as nodes and edges as similarity between them
- TextRank (Mihalcea & Tarau, 2004)
  - Similarity defined as the number of similar words

$$Similarity(S_i, S_j) = \frac{|\{w_k | w_k \in S_i \& w_k \in S_j\}|}{log(|S_i|) + log(|S_j|)}$$

  - Vertex scores calculated using PageRank

$$S(V_i) = (1-d) + d * \sum_{j \in In(V_i)} \frac{1}{|Out(V_j)|} S(V_j)$$

- LexRank (Erkan & Radev, 2004)
  - Similarity as cosine similarity between sentences



Content Selection Methods (Discourse-based)

- Rhetorical Structure Theory (Mann & Thomson, 1988)
- Focus on coherence

  - I love pets. They make me happy.
    Nucleus Satellite

[The truth is that the pressure to smoke in junior high is greater than it will be any other time of one's life:]$_N$ [we know that 3,000 teens start smoking each day.]$_S$

Improving summarization via Rhetorical Parsing (Marcu 1998)

Sentence Selection

Select important sentences to form a paragraph length summary

- Choose desired summary length
  - Pick sentences
- How to pick sentences?
  - Sentences should be relevant and non-redundant
- Maximal marginal relevance approach (Carbonell & Goldstein 1998)
  - Iterative greedy procedure
  - at each step compute sentence importance score
    - linear combination of previous importance weight and similarity with already chosen sentences

Summarization Tasks

- Content selection
  - Choose sentences to extract
- Information ordering
  - Order sentences
- Realization
  - Cleanup and present

Sentence Realization

- Sentences selected and ordered
- Need sentence compression or simplification
  - eliminate adjective modifiers and subordinate clauses

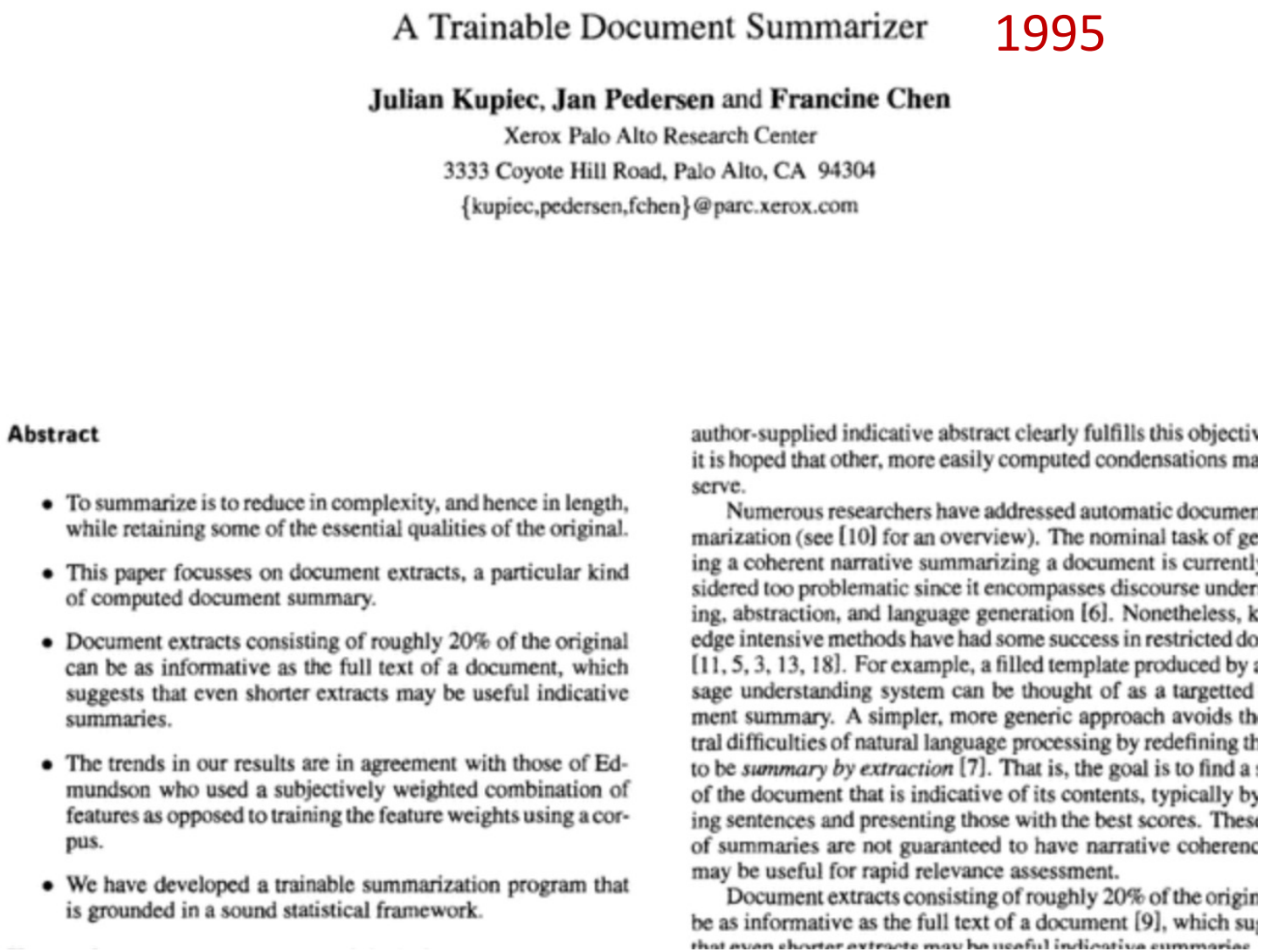
(23.31) **Original sentence:** ~~When it arrives sometime new year in new TV sets,~~ the V-chip will give parents a ~~new and potentially revolutionary~~ device to block out programs they don't want their children to see.

(23.32) **Simplified sentence by humans:** The V-chip will give parents a device to block out programs they don't want their children to see.

Supervised Methods

- Given:
  - a labeled training set of good summaries for each document
- Align:
  - the sentences in the document with sentences in the summary
- Extract features
  - position (first sentence?)
  - length of sentence
  - word informativeness, cue phrases
  - cohesion
- Train
  - a binary classifier
    (put sentence in summary? yes or no)
- Problems:
  - hard to get labeled training
  - alignment difficult
  - performance not better than unsupervised algorithms
- So in practice:
  - Unsupervised content selection more common

Supervised Methods



A Trainable Document Summarizer 1995

Julian Kupiec, Jan Pedersen and **Francine Chen**
Xerox Palo Alto Research Center
3333 Coyote Hill Road, Palo Alto, CA 94304
{kupiec,pedersen,fchen}@parc.xerox.com

Abstract

- To summarize is to reduce in complexity, and hence in length, while retaining some of the essential qualities of the original.
- This paper focusses on document extracts, a particular kind of computed document summary.
- Document extracts consisting of roughly 20% of the original can be as informative as the full text of a document, which suggests that even shorter extracts may be useful indicative summaries.
- The trends in our results are in agreement with those of Edmundson who used a subjectively weighted combination of features as opposed to training the feature weights using a corpus.
- We have developed a trainable summarization program that is grounded in a sound statistical framework.

author-supplied indicative abstract clearly fulfills this objecti
it is hoped that other, more easily computed condensations ma
serve.

Numerous researchers have addressed automatic documer
marization (see [10] for an overview). The nominal task of ge
ing a coherent narrative summarizing a document is currentl
sidered too problematic since it encompasses discourse under
ing, abstraction, and language generation [6]. Nonetheless, k
edge intensive methods have had some success in restricted do
[11, 5, 3, 13, 18]. For example, a filled template produced by
sage understanding system can be thought of as a targetted
ment summary. A simpler, more generic approach avoids th
tral difficulties of natural language processing by redefining th
to be *summary by extraction* [7]. That is, the goal is to find a
of the document that is indicative of its contents, typically by
ing sentences and presenting those with the best scores. These
of summaries are not guaranteed to have narrative coherenc
may be useful for rapid relevance assessment.

Document extracts consisting of roughly 20% of the origi
be as informative as the full text of a document [9], which su

- Naïve Bayes
- Sentence length
- Set of phrases
- Sentence position
- P(s|features)

Extractive vs Abstractive

- Extractive simpler and more accurate
- But, abstractive more human-like

- To improve in extractive
  - Redundancy
  - Temporal ordering for multi-document

Extractive vs Abstractive

- Extractive simpler and more accurate
- But, abstractive more human-like

- To improve in extractive
  - Redundancy
  - Temporal ordering for multi-document

Abstractive Summarization

Source: the sri lanka government on wednesday announced the closure of government schools with immediate effect as a military campaign against tamil separatists escalated in the north of the country.
Summary: sri lanka **closes** schools as **war escalates**.

- Copying chunks of text from the source ensures baseline levels of grammaticality and accuracy
- High-quality summarization relies on paraphrasing, generalization

Abstractive Summarization

- Sentence Compression (Cohn & Lapata 2008), sentence fusion (Barzilay & McKeown 2005)
  - Paraphrasing and new text generation
  - Relies on syntactic tree rewriting via a set of rules + LM

Information Extraction

- Information Extraction
  - Extract subject-verb-subject triple
- Content Selection
  - Select subset of candidates subject to length constraints
  - Solved using Integer Linear Programming (Murray et al 2010, Woodsend & Lapata 2011, Bing et al. 2015)
    - Optimize an objective function (weighted sum of a set of binary variables) subject to a set of linear constraints
    - Weight associated with variable indicates importance of candidate phrase
    - Phrases learned jointly
    - Constraints such as length

Graph-Based Summarization

- *Event semantic link networks* (ESLNs) (Li et al. 2016).
  - Node: event mentioned in input text, event is event trigger/action and its arguments.
  - Edge between two nodes: Semantic relation between corresponding events
  - ILP applied for information extraction and content selection (i.e., selecting a subset of nodes for generating the summary)
    - use length and semantic relations (e.g., the nodes should be chosen such that the resulting graph remains connected) constraints

Template-Based Summarization

- Human summaries (e.g., meeting summaries for accomplishing a certain task) have common sentence structures
  - learned from human summaries in training set (*templates)*
- Summary generated by learning and then filling template (Oya et al, 2014)
  - Template learning
    - Replace each sentence NP with hypernym
    - Cluster sentences based on their root verbs
    - Generate representative sentence templates by graph-based method
  - Keyphrase extraction (label phrases with hypernym)
  - templates with highest similarity with each topic segment of meeting are selected
  - Sentence ranker to select sentences for summary

Neural Methods

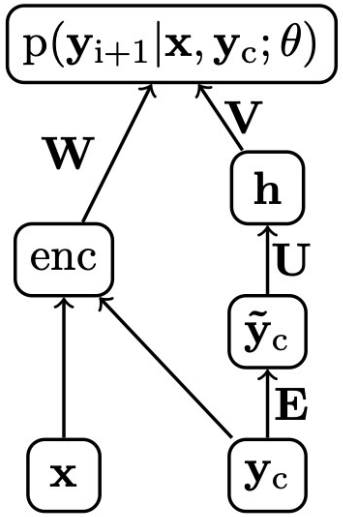
- End-to-end approach alternative to information extraction, content selection and surface realization

- Less control over what is learned and how information is encoded

Neural Methods

- Extractive summarization
  - word or sentence level classification problem
  - Solved using representations (Cheng and Lapata 2016; Nallapati et al. 2017; Xu and Durrett 2019)

- Abstractive summarization
  - NNLM-based (Rush et al. 2015); transformer-based (Duan et al. 2019)



$$
\begin{aligned}
p(\mathbf{y}_{i+1}|\mathbf{y}_c, \mathbf{x}; \theta) &\propto \exp(\mathbf{V}\mathbf{h} + \mathbf{W}\mathrm{enc}(\mathbf{x}, \mathbf{y}_c)), \\
\tilde{\mathbf{y}}_c &= [\mathbf{E}\mathbf{y}_{i-C+1}, \ldots, \mathbf{E}\mathbf{y}_i], \\
\mathbf{h} &= \tanh(\mathbf{U}\tilde{\mathbf{y}}_c).
\end{aligned}
$$

Neural Methods

- Encoder performs information extraction
  - Encoding long documents made easier by extractive summarization (Chen & Bansal, 2018)
  - Remains a challenge

- Exploiting background knowledge (e.g., about entities) to aid decoder (Amplayo et al. 2018)

- Redundancy problem addressed
  - Distraction (Nema et al 2017)
  - Coverage loss (see et al 2017) for repetition

Evaluation

- Summary length
- Fidelity
- Grammatical
- Non-redundant
- Referentially well-formed
- Coherent

Evaluation

- Extrinsic
  - Task-based
  - Can you make the same decision using summary as with full text? Less time?

- Intrinsic
  - Compare generated summary with gold summary

Evaluation

- Recall Oriented Understudy for Gisting Evaluation (ROUGE; Lin and Hovy 2003)
  - Measures *N*-gram overlap between candidate and human-generated summaries (the references)
  - ROUGE-N

$$ROUGE2 = \frac{\sum_{S \in \{ReferenceSummaries\}} \sum_{bigram \in S} \text{Count}_{\text{match}}(bigram)}{\sum_{S \in \{ReferenceSummaries\}} \sum_{bigram \in S} \text{Count}(bigram)}$$

  - ROUGE-L
    - Longest common subsequence instead of n-gram

Evaluation

Limitations of automatic metrics

- Limited correlation with human judgments
- No measures for factual consistency

Datasets

- Document Understanding Conference and Text Analysis Conference
  - English News articles
  - Generic and focused summarization
  - Small (a few hundred)

- Annotated English Gigaword
  - ~10 million documents
  - First sentence of source (text), headline (summary)

- CNN/Daily Mail
  - ~300K documents
  - Multi-sentence summaries

Neural Methods

Methods / ROUGE	Extractive Methods				Abstractive Methods					
	Lead-3	**TextRank**	**Summa**	**BertExt**	**S2S**	**PG**	**PG-Coverage**	**Bottom-Up**	**BertAbs**	**BART**
ROUGE-1	39.20	40.20	39.60	43.25	31.33	36.44	39.53	41.22	42.13	**44.16**
ROUGE-2	15.70	17.56	16.20	20.24	11.81	15.66	17.28	18.68	19.60	**21.28**
ROUGE-L	35.50	36.44	35.30	39.63	28.80	33.42	36.38	38.34	39.18	**40.90**

Table 1: ROUGE scores of 10 summarizers on CNN/DM Dataset (non-anonymous version). We get the score of Lead-3 and TextRank from Nallapati et al. (2017) and Zhou et al. (2018), respectively.

Take-aways

- Extractive summarizers better than their abstractive counterparts
  - strength in faithfulness and factual-consistency

- Techniques such as copy, coverage and hybrid extractive/abstractive methods bring specific improvements but also demonstrate limitations

- Pre-training techniques, particular sequence-to-sequence pre-training, highly effective for summarization

Challenges

- *Accuracy*-related: summary does not reflect the source
  - Addition, Omission, Inaccuracy, Positive-Negative Aspect

- *Fluency* issues refer to linguistic qualities of the text.