From Standard Summarization to New Tasks and Beyond: Summarization with Manifold Information

Shen Gao and Rui Yan
shengao@pku.edu.cn
Outline

1. Introduction to summarization with manifold information
   1. Traditional summarization
   2. New summarization task
   3. Challenges and problems
2. Background knowledge
   1. Concepts, problem formulation, and task statements
   2. Deep learning for summarization
3. Summarization by incorporating document structure
   1. Long document summarization
   2. Timeline summarization
   3. Dialog summarization
   4. Academic paper summarization
4. Summarization by incorporating additional knowledge
   1. Reader-aware summarization
   2. Template-based summarization
   3. Multi-modal summarization
   4. Opinion Summarization
5. Recent trends
   1. Multi-modal summarization
   2. Long document summarization
   3. Dialog summarization
6. Summary

5. Movie Summarization
Target Audience

• Our target audiences are researchers and practitioners with some deep learning and text process background

• Our target audiences are interested in new summarization task and the technologies behind the prosperity of real-world summarization application in industry and academia.

• They would like to learn how to build a summarization system with state-of-the-art technologies.
Introduction

• Two types of text summarization
  • Summarizes a plain text
  • Generating summary with manifold information

• New summarization tasks aim to produce a better and appropriate summary by incorporating manifold information in many real-world applications.
Task of Traditional summarization

• Very simple and general
• Input: a plain text document
• Output: a short dense text describe the main idea of the input document
New summarization task

• Different with traditional summarization task
• Using structured document as input
• Leveraging other knowledge source as additional input
• These new summarization task can better adapt to real-world summarization applications
Challenges and problems

• How to understand the semantic meanings of the text with structure?
• How to incorporate additional knowledge when summarizing documents?
Background: Deep Learning for Summarization

• Extractive Summarization
  • **Sequence Labeling** uses an RNN to read the sentences only once
  • **Encoder-Decoder** uses two RNN to encode the passage and decode the sentence pointer.
  • **Reinforcement learning** method directly optimize the ROUGE score
  • **Pretraining** techniques employ the language model pre-training model
  • **Graph Model** contains additional nodes which act as the intermediary between sentences and enrich the cross-sentence relations

• Abstractive Summarization
  • **Sequence-to-sequence** based text generation methods
  • **Copy mechanism** directly copy the OOV words
  • **Selective encoding** encode the important semantic parts and ignore the trivial parts.
  • **Pretraining** techniques employ the language model pre-training model
  • **Contrastive Learning** bridge the gap between the *learning objective* and *evaluation metrics*
Datasets

• CNNDM
• WikiSum
• BIGPATENT
• Newsroom
• WikiHow
• XSUM
Extractive Summarization

**Gold Summary:**
Redpath has ended his eight-year association with Sale Sharks. Redpath spent five years as a player and three as a coach at Sale. He has thanked the owners, coaches and players for their support.

<table>
<thead>
<tr>
<th>Sentence</th>
<th>Salience</th>
<th>Content</th>
<th>Novelty</th>
<th>Position</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bryan Redpath has left his coaching role at Sale Sharks with immediate effect.</td>
<td>0.1</td>
<td>0.1</td>
<td>0.9</td>
<td>0.1</td>
<td>0.3</td>
</tr>
<tr>
<td>The 43-year-old Scot ends an eight-year association with the Aviva Premiership side, having spent five years with them as a player and three as a coach.</td>
<td>0.9</td>
<td>0.6</td>
<td>0.9</td>
<td>0.9</td>
<td>0.7</td>
</tr>
<tr>
<td>Redpath returned to Sale in June 2012 as director of rugby after starting a coaching career at Gloucester and progressing to the top job at Kingsholm.</td>
<td>0.8</td>
<td>0.5</td>
<td>0.5</td>
<td>0.9</td>
<td>0.6</td>
</tr>
<tr>
<td>Redpath spent five years with Sale Sharks as a player and a further three as a coach but with Sale Sharks struggling four months into Redpath’s tenure, he was removed from the director of rugby role at the Salford-based side and has since been operating as head coach.</td>
<td>0.8</td>
<td>0.9</td>
<td>0.7</td>
<td>0.8</td>
<td>0.9</td>
</tr>
<tr>
<td>‘I would like to thank the owners, coaches, players and staff for all their help and support since I returned to the club in 2012.</td>
<td>0.4</td>
<td>0.1</td>
<td>0.1</td>
<td>0.7</td>
<td>0.2</td>
</tr>
<tr>
<td>Also to the supporters who have been great with me both as a player and as a coach,’ Redpath said.</td>
<td>0.6</td>
<td>0.0</td>
<td>0.2</td>
<td>0.3</td>
<td>0.2</td>
</tr>
</tbody>
</table>
Extractive Summarization

• Composed of a hierarchical document encoder and an attention-based extractor.
• Reader is to derive the meaning representation of a document based on its sentences and their constituent words.
• Sentence extractor applies attention to directly extract salient sentences after reading them.
Extractive Summarization

• Human usually select salient sentences and then rewrite them as the final summary.
• Sentence-level policy gradient method to bridge the non-differentiable computation between these two neural networks in a hierarchical way.
Extractive Summarization

• Language model pretraining has advanced the state of the art in many NLP tasks
• Explore the potential of BERT for text summarization under a general framework
• Experiments on three datasets show that this model achieves state-of-the-art results
Extractive Summarization

• Contains semantic nodes of different granularity levels apart from sentences
• These additional nodes act as the intermediary between sentences and enrich the cross-sentence relations.
Abstractive Summarization

• Sequence-to-sequence models have provided a viable new approach for *abstractive* text summarization

• A hybrid pointer-generator network that can copy words from the source text via *pointing*, which aids accurate reproduction of information
Abstractive Summarization

• A unified model to combine the strength of both state-of-the-art extractor and abstracter.

• Inconsistency loss function is introduced to penalize the inconsistency between two levels of attentions.
Abstractive Summarization

- Pre-training Transformers with self-supervised objectives on large text corpora has shown great success when fine-tuned on downstream NLP tasks including text summarization
- Important sentences are removed/masked from an input document and are generated together as one output sequence from the remaining sentences

<table>
<thead>
<tr>
<th>R1/R2/RL</th>
<th>XSum</th>
<th>CNN/DailyMail</th>
<th>Gigaword</th>
</tr>
</thead>
<tbody>
<tr>
<td>BERTShare (Rothe et al., 2019)</td>
<td>38.52/16.12/31.13</td>
<td>39.25/18.09/36.45</td>
<td>38.13/19.81/35.62</td>
</tr>
<tr>
<td>MASS (Song et al., 2019)</td>
<td>39.75/17.24/31.95</td>
<td>42.12/19.50/39.01</td>
<td>38.73/19.71/35.96</td>
</tr>
<tr>
<td>UniLM (Dong et al., 2019)</td>
<td>-</td>
<td>43.33/20.21/40.51</td>
<td>38.45/19.45/35.75</td>
</tr>
<tr>
<td>BART (Lewis et al., 2019)</td>
<td>45.14/22.27/37.25</td>
<td>44.16/21.28/40.90</td>
<td>-</td>
</tr>
<tr>
<td>T5 (Raffel et al., 2019)</td>
<td>-</td>
<td>43.52/21.55/40.69</td>
<td>-</td>
</tr>
<tr>
<td>PEGASUS\textsubscript{LARGE} (C4)</td>
<td>45.20/22.06/36.99</td>
<td>43.90/21.20/40.76</td>
<td>38.75/19.96/36.14</td>
</tr>
<tr>
<td>PEGASUS\textsubscript{LARGE} (HugeNews)</td>
<td>47.21/24.56/39.25</td>
<td>44.17/21.47/41.11</td>
<td>39.12/19.86/36.24</td>
</tr>
</tbody>
</table>
Incorporating Document Structure

• Timeline Summarization
  • help users to have a quick understanding of the overall evolution of any given topic
  • consider evolutionary characteristics of news plus to traditional summary elements

• Extreme Long Document Summarization
  • the input document can be very long, such as an academic paper or a patent document which is longer than the news article
  • extract the salient information and central idea from a large amount of information.

• Dialog Summarization
  • time-consuming for people to review all the context before starting a new dialog
  • the salient information is scattered in the whole dialog history

• Academic paper summarization
  • The reference relationship should be considered into generating summary of academic paper.

• Movie Summarization
  • Summarizing longer narratives, screenplays, whose form and structure is far removed from newspaper articles.
Timeline Summarization

• Timeline summarization is an important research task which can help users to have a quick understanding of the overall evolution of any given topic.
• The previous works are all based on extractive methods
• A large-scale dataset with 169,423 training samples, 5,000 evaluation and 5,000 test samples.
• On average, there are 352.22 words and 61.16 words in article and summary respectively.

Learning towards Abstractive Timeline Summarization
Timeline Summarization

- Given any collection of time-stamped news articles, MTLS automatically discovers important yet different stories and generates a corresponding timeline for each story.
- Propose a Two-Stage Affinity Propagation Summarization framework which is a two-stage clustering-based framework.

<table>
<thead>
<tr>
<th>MTLS Methods</th>
<th><strong>concat</strong></th>
<th><strong>align+m:1</strong></th>
<th><strong>agreement</strong></th>
<th><strong>d-select</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ROUGE-1</td>
<td>ROUGE-2</td>
<td>ROUGE-1</td>
<td>ROUGE-2</td>
</tr>
<tr>
<td>Baselines</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHIEU/2004</td>
<td>Random</td>
<td>0.191</td>
<td>0.027</td>
<td>0.019</td>
</tr>
<tr>
<td></td>
<td>LDA</td>
<td>0.192</td>
<td>0.035</td>
<td>0.023</td>
</tr>
<tr>
<td></td>
<td>k-means</td>
<td>0.229</td>
<td>0.046</td>
<td>0.027</td>
</tr>
<tr>
<td>MARTSCHAT2018</td>
<td>Random</td>
<td>0.254</td>
<td>0.049</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td>LDA</td>
<td>0.289</td>
<td>0.068</td>
<td>0.062</td>
</tr>
<tr>
<td></td>
<td>k-means</td>
<td>0.291</td>
<td>0.071</td>
<td>0.061</td>
</tr>
<tr>
<td>GHALANDARI2020</td>
<td>Random</td>
<td>0.253</td>
<td>0.048</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td>LDA</td>
<td>0.268</td>
<td>0.062</td>
<td>0.085</td>
</tr>
<tr>
<td></td>
<td>k-means</td>
<td>0.284</td>
<td>0.073</td>
<td><strong>0.096</strong></td>
</tr>
<tr>
<td><strong>Our method</strong></td>
<td>2SAPS</td>
<td><strong>0.312</strong></td>
<td><strong>0.084</strong></td>
<td><strong>0.096</strong></td>
</tr>
</tbody>
</table>

Multi-TimeLine Summarization (MTLS): Improving Timeline Summarization by Generating Multiple Summaries
Extreme Long Document Summarization

• Datasets of long document summarization task
• A hierarchical encoder, capturing the discourse structure of the document.
• A discourse-aware decoder that generates the summary.

<table>
<thead>
<tr>
<th>Dataset</th>
<th># Doc</th>
<th>Summary # word</th>
<th>Doc # sent</th>
<th>Comp. Den.</th>
<th># word</th>
</tr>
</thead>
<tbody>
<tr>
<td>PUBMED</td>
<td>133,215</td>
<td>202.4</td>
<td>6.8</td>
<td>3049.0</td>
<td>16.2</td>
</tr>
<tr>
<td>ARXIV</td>
<td>215,913</td>
<td>272.7</td>
<td>9.6</td>
<td>6029.9</td>
<td>39.8</td>
</tr>
<tr>
<td>BILLSUM</td>
<td>23,455</td>
<td>207.7</td>
<td>7.2</td>
<td>1813.0</td>
<td>13.6</td>
</tr>
<tr>
<td>BIGPATENT</td>
<td>1,341,362</td>
<td>116.5</td>
<td>3.7</td>
<td>3573.2</td>
<td>36.3</td>
</tr>
<tr>
<td>GOVEREPORT</td>
<td>19,466</td>
<td>553.4</td>
<td>17.8</td>
<td>9409.4</td>
<td>19.0</td>
</tr>
</tbody>
</table>

A Discourse-Aware Attention Model for Abstractive Summarization of Long Documents
Extreme Long Document Summarization

- Sliding selector network with dynamic memory for extractive summarization of long-form documents
- A memory to preserve salient information learned from previous windows
Extreme Long Document Summarization

• The main challenge of summarizing long document is how to find salient information from large amount of sentences effectively.

• Encoder-decoder attention with head-wise positional strides
## Dialog Summarization - Datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>MEDIA$\text{SUM}$</th>
<th>AMI</th>
<th>ICSI</th>
<th>DiDi</th>
<th>CRD3</th>
<th>MultiWOZ</th>
<th>SAM$\text{SUM}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
<td>Transcribed Speech</td>
<td>Written</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type</td>
<td>Interview</td>
<td>Meeting</td>
<td>Meeting</td>
<td>Customer</td>
<td>Game</td>
<td>Booking</td>
<td>Daily</td>
</tr>
<tr>
<td>Real dialogue</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Open domain</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>✓</td>
<td>×</td>
<td>✓</td>
</tr>
<tr>
<td>Public</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Dialogues</td>
<td>463,596</td>
<td>137</td>
<td>59</td>
<td>328,880</td>
<td>159</td>
<td>10,438</td>
<td>16,369</td>
</tr>
<tr>
<td>Dial. words</td>
<td>1,553.7</td>
<td>4,757</td>
<td>10,189</td>
<td>/</td>
<td>31,802.8</td>
<td>180.7</td>
<td>83.9</td>
</tr>
<tr>
<td>Summ. words</td>
<td>14.4</td>
<td>322</td>
<td>534</td>
<td>/</td>
<td>2062.3</td>
<td>91.9</td>
<td>20.3</td>
</tr>
<tr>
<td>Turns</td>
<td>30.0</td>
<td>289</td>
<td>464</td>
<td>/</td>
<td>2,507.4</td>
<td>13.7</td>
<td>9.9</td>
</tr>
<tr>
<td>Speakers</td>
<td>6.5</td>
<td>4</td>
<td>6.2</td>
<td>2</td>
<td>9.6</td>
<td>2</td>
<td>2.2</td>
</tr>
</tbody>
</table>

**MEDIA$\text{SUM}$**: A Large-scale Media Interview Dataset for Dialogue Summarization
Dialog Summarization

• In this section we describe the complete pipeline of the model which includes (1) Sequence labelling of utterance tags, (2) Re-ordering of conversation to model discourse relations, and (3) Pointer-generator, coverage based model for abstractive summarization.
Dialog Summarization

• A meeting is naturally full of dialogue-specific structural information

• Previous works model a meeting in a sequential manner, while ignoring the rich structural information

• Dialogue discourse is a dialogue-specific structure that can provide pre-defined semantic relationships between each utterance
Dialog Summarization

- **Topic View**: Based on what topics were discussed, it can be segmented into several topics.
- **Stage View**: From a conversation progression perspective.
- **Global View**: Conversations can be treated as a whole.
- **Discrete View**: Each utterance can serve as one segment.
Dialog Summarization

• Existing features are obtained via open-domain toolkits that are dialog-agnostic or heavily relied on human annotations

• Perform three dialogue annotation tasks takes advantage of dialogue background knowledge encoded in DialoGPT

Language Model as an Annotator: Exploring DialoGPT for Dialogue Summarization
Dialog Summarization

• Existing generated dialog summaries often suffer from insufficient, redundant, or incorrect content.
• Explicitly model the rich structures in conversations for more precise and accurate conversation summarization.

Structure-Aware Abstractive Conversation Summarization via Discourse and Action Graphs
Topic-Oriented Spoken Dialogue Summarization for Customer Service with Saliency-Aware Topic Modeling
## Academic paper summarization

- **Datasets**

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Source</th>
<th># Pairs</th>
<th>Doc. Length</th>
<th>Sum. Length</th>
<th># Sections</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Train</td>
<td># Words</td>
<td># Words</td>
<td></td>
</tr>
<tr>
<td>CNN</td>
<td>News</td>
<td>90,266</td>
<td>760.5</td>
<td>45.7</td>
<td>-</td>
</tr>
<tr>
<td>DailyMail</td>
<td>News</td>
<td>196,961</td>
<td>653.3</td>
<td>54.7</td>
<td>-</td>
</tr>
<tr>
<td>ScisummNet</td>
<td>Scientific Papers</td>
<td>1009</td>
<td>4203.4</td>
<td>150.7</td>
<td>6.5</td>
</tr>
<tr>
<td>arXiv†</td>
<td>Scientific Papers</td>
<td>215,913</td>
<td>4938.0</td>
<td>220.0</td>
<td>5.9</td>
</tr>
<tr>
<td>PubMed‡</td>
<td>Scientific Papers</td>
<td>119,924</td>
<td>3016.0</td>
<td>203.0</td>
<td>5.6</td>
</tr>
<tr>
<td>SSN (inductive)</td>
<td>Scientific Papers</td>
<td>128,400</td>
<td>5072.3</td>
<td>165.1</td>
<td>10.8</td>
</tr>
<tr>
<td>SSN (transductive)</td>
<td>Scientific Papers</td>
<td>128,299</td>
<td>5072.3</td>
<td>165.1</td>
<td>10.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Val</td>
<td></td>
<td># Sent.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1,220</td>
<td>34.0</td>
<td>3.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>12,148</td>
<td>29.3</td>
<td>3.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>6440</td>
<td>206.3</td>
<td>9.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>6633</td>
<td>86.4</td>
<td>6.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>6123</td>
<td>290.6</td>
<td>6.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>6250</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Academic paper summarization

- Integrating the authors’ original highlights (abstract) and the article’s actual impacts on the community

ScisummNet: A Large Annotated Corpus and Content-Impact Models for Scientific Paper Summarization with Citation Networks
Academic paper summarization

Enhancing Scientific Papers Summarization with Citation Graph
Academic paper summarization

• Given a set of related publications, related work section generation aims to provide researchers with an overview of the specific research area by summarizing these works and introducing them in a logical order.
Movie Summarization

- Most efforts to date have concentrated on the summarization of news articles.
- Screenplays, whose form and structure is far removed from newspaper articles.
Incorporating Additional Knowledge

- Reader-aware Summarization
  - reader comments concentrate on the main idea of the news article
  - comments can be used to help the summarization model to capture the main idea

- Template Based Summarization
  - first retrieves a summary template and then edits it into the new summary of the current document.

- Multi-Modal Summarization
  - increase of multi-media data on the web
  - the visual information is incorporated along with the input document into the text summarizing process to improve the quality

- Query-based Summarization
  - search engine provides a list of web pages associated with their summaries
  - should summarize the query focused aspect of the web page instead of the main idea

- Opinion Summarization
  - Summarize the opinion of e-commerce reviews
Reader-aware Summarization

• In the beginning, researchers firstly propose to understand the input document with the feedback of readers using a graph-based method, where they identify three relations (topic, quotation, and mention) by which comments can be linked to one another.
Reader-aware Summarization

Employ a sparse coding based framework for this task which jointly considers news documents and reader comments via an unsupervised data reconstruction strategy.
Reader-aware Summarization

- A large-scale reader-aware summarization dataset (863826 training samples)
- A generative-adversarial based method which conducts the interaction between reader comments and news to capture the reader attention distribution on the article
Template Based Summarization

• Previous seq2seq purely depend on the source, which tends to work unstably
• Use a popular IR platform to Retrieve proper summaries as candidate templates
• Extend the seq2seq framework to jointly conduct template Reranking and template-aware summary generation
Template Based Summarization

• Bi-directional Selective Encoding with Template (BiSET) model
• Leverages template discovered from training data to softly select key information from each source article
• A multi-stage process for automatic retrieval of high-quality templates from training corpus.
Template Based Summarization

• In circumstances, the generated summaries are required to conform to a specific pattern.
• Template-based methods are too rigid for our patternized summary generation task.
• We propose a summarization framework named Prototype Editing based Summary Generator that incorporates prototype document-summary pairs.

How to Write Summaries with Patterns? Learning towards Abstractive Summarization through Prototype Editing
Multi-Modal Summarization - Image

• Multimodal Summarization with Multimodal Output

• Four modules: text encoder, image encoder, multimodal attention layer, and summary decoder

• Propose a multimodal automatic evaluation (MMAE) method which mainly considers three aspects: salience of text, salience of image, and relevance between text and image.

MSMO: Multimodal Summarization with Multimodal Output
Multi-Modal Summarization

- Existing MSMO methods are trained by the target of text modality
- Leading to the modality-bias problem
- Propose a multimodal objective function with the guidance of multimodal reference
Multi-Modal Summarization - Video

- Video and document as input
- Selects cover frame from news video and generates textual summary of the news article in the meantime

VMSMO: Learning to Generate Multimodal Summary for Video-based News Articles
Multi-Modal Summarization – Ecommerce

- Adopt an aspect-based reward augmented maximum likelihood training method
- Aspect coverage mechanism to keep track of what aspects have been mentioned
- Adopt constrained decoding to enhance the coherence of summaries
Query-based Summarization

- Query-based summarization highlights those points that are relevant to the user query
- Seq2seq suffers from the drawback of generation of repeated phrases
- A query attention model which learns to focus on different portions of the query
- A new diversity based attention model

Diversity driven attention model for query-based abstractive summarization
Query-based Summarization

- A coarse-to-fine modeling framework for extractive query focused summarization which incorporates a relevance estimator, an evidence estimator and a centrality estimator.
Query-based Summarization

- Existing methods are limited by the lack of sufficient large-scale high-quality training datasets.
- Present two QMDS training datasets: (1) QMDSCNN and (2) QMDSIR by using two data augmentation methods.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Train</th>
<th>Val</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>QMDSCNN (# samples)</td>
<td>287,113</td>
<td>13,368</td>
<td>11,490</td>
</tr>
<tr>
<td>- Avg. # documents</td>
<td>6.5</td>
<td>6.5</td>
<td>6.5</td>
</tr>
<tr>
<td>- Avg. Doc. length (# tokens)</td>
<td>355</td>
<td>346</td>
<td>353</td>
</tr>
<tr>
<td>- Avg. Query length (# tokens)</td>
<td>13.8</td>
<td>14.5</td>
<td>14.2</td>
</tr>
<tr>
<td>QMDSIR (# samples)</td>
<td>82,076</td>
<td>10,259</td>
<td>10,260</td>
</tr>
<tr>
<td>- Avg. # documents</td>
<td>5.8</td>
<td>5.4</td>
<td>5.5</td>
</tr>
<tr>
<td>- Avg. Doc. length (# tokens)</td>
<td>1,291</td>
<td>1,402</td>
<td>1,379</td>
</tr>
<tr>
<td>- Avg. Query length (# tokens)</td>
<td>6.2</td>
<td>6.2</td>
<td>6.2</td>
</tr>
</tbody>
</table>
Query-based Summarization

- Hierarchical query focused order-aware multi-document summarization model:
  - Hierarchical Encoding
  - Ordering Component
  - Query Component

Data Augmentation for Abstractive Query-Focused Multi-Document Summarization
Opinion Summarization

• Abstractive opinion summarization framework, which does not rely on gold-standard summaries for training
• Uses an Aspect-based Sentiment Analysis model to extract opinion phrases from reviews, and trains a Transformer model to reconstruct the original reviews from these extractions
Opinion Summarization

- Training data is not available and cannot be easily sourced
- Create a synthetic dataset from a corpus of user reviews by sampling a review, pretending it is a summary
- Generating noisy versions thereof which we treat as pseudo-review input
Opinion Summarization

• Training data is neither available nor can be easily sourced
• Explicitly incorporating *content planning* in a summarization model allows the creation of synthetic datasets

Unsupervised Opinion Summarization with Content Planning
Recent trends

- Multi-modal summarization
- Long document summarization
- Dialog summarization
Thanks!

Email: shengao@pku.edu.cn