The Initial End-to-end Summarization System

Long Cheng, Sheng Bi, Catherine Wang, Vicky Xiang, Carrie Yuan
University of Washington

System architecture

Content Selection → Information Ordering → Content Realization

Content Selection

The first stage of the system is Content Selection, where various algorithms are applied to identify the most important information from the input text.

Three algorithms, namely LLR, BasicSum, and LexRank, are employed for Content Selection. These algorithms analyze the text and prioritize sentences or phrases based on their relevance, importance, and semantic connections.

The output of Content Selection is a set of selected sentences or phrases deemed most crucial for inclusion in the summary.

Each algorithm's performance is evaluated using ROUGE scores, providing a quantitative measure of how well the selected content captures the essence of the original text.

Information Ordering

The next stage of the system, Information Ordering, is responsible for structuring the selected content in a logical and coherent manner.

This component arranges the selected sentences or phrases in a sequence that enhances the readability and comprehension of the summary.

Information Ordering ensures that the summary flows smoothly and effectively communicates the key points of the original text.

Content Realization

The final stage of the system, Content Realization, focuses on generating the actual summary based on the ordered content.

This component may involve text generation techniques such as sentence fusion, paraphrasing, or abstraction to create a concise and informative summary.

Content Realization aims to produce summaries that capture the essence of the original text while minimizing redundancy and irrelevant details.

Approaches

Content Selection

Log Likelihood Ratio (LLR)

- k_1= count of w in the input documents
- k_2= count of w in background corpus
- n_1= total # of words in the input documents
- n_2= total # of words in the background corpus
- $p_1=k_1/n_1$; $p_2=k_2/n_{2;}$ $p= (k_1+k_2)/(n_1+n_2)$

- Binomial distribution: $L(p, k, n) = C(n, k)\, p^k\, (1 - p)^{n-k.}$
- H0: P(w)=p for both input and background corpus
- H1: P(w)=p1 for the input, P(w)=p2 for the background corpus
- $\lambda = L(X \mid H0)/L(X \mid H1)$
  $= L(p, k_1, n_1)L(p, k_2, n_2)/L(p_1, k_1, n_1)L(p_2, k_2, n_2)$

$$-2log\lambda = 2[logL(p_1, k_1, n_1) + logL(p_2, k_2, n_2)$$
$$-logL(p, k_1, n_1) - logL(p, k_2, n_2)]$$

Some good examples of LLR

```
mad cow disease , or bovine spongiform encephalopathy , eats holes in the brains of cattle .
the likely vector of contamination for livestock was brain and nerve tissue mixed in animal feed .
ireland banned the use of meat and bone meal as cattle feed , the suspected origin of mad cow disease , in 1990 .
the fatal brain-wasting disease is believed to come from eating beef products from cows struck with mad cow disease .
the vast bulk of them are elderly dairy cattle who would have eaten cattle-based feed in the 1980s .
```

```
george bush says we can wait .
carolyn susman writes for the palm beach post .
`` john kerry strongly supports stem cell research .
symptoms can be as diverse as loss of spontaneous movement , rigidity , tremor and shuffling gait , acute confusion , memory loss and problems with other mental functions .
the loss of cells that produce the neurotransmitter dopamine causes the telltale tremors , rigid and slow movements of parkinson 's .
three years ago , bush limited federal funding of embryonic stem cell research to the 78 stem cell lines in existence .
```

```
a california laboratory reportedly digitally enhanced snippets of conversation picked up at 5:52 a.m. from the ramsey kitchen .
newsweek reporters sherry keene-osborn and daniel glick , who are associate producers of the documentary `` who killed jonbenet ?
but susan gordy of atlanta said the ramseys ' stated reason for waiting makes sense .
in two separate letters , they accuse boulder district attorney alex hunter of unwarranted delays in prosecuting the case and of being overly accommodating to the ramseys .
the ramseys have proclaimed their innocence .
```

Some issues of LLR

Very similar sentences were selected.

```
twenty-nine of 71 potential jurors have been dismissed .
twenty-two of 71 potential jurors have been dismissed .
```

Not complete English sentences were selected.

```
attention - updates toll, adds rail blast ///
```

```
3 , box 260 , ridgeland , s.c. 29936 .
```

```
undated : dui vs. dwi ?
```

SumBasic

The basic algorithm

Step 1 Compute the probability distribution over the words w_i appearing in the input, $p(w_i)$ for every i; $p(w_i) = \frac{n}{N}$, where n is the number of times the word appeared in the input, and N is the total number of content word tokens in the input.

Step 2 For each sentence S_j in the input, assign a weight equal to the average probability of the words in the sentence, i.e.

$weight(S_j) = \sum_{w_i \in S_j} \frac{p(w_i)}{|\{w_i | w_i \in S_j\}|}$

Step 3 Pick the best scoring sentence that contains the highest probability word.

Step 4 For each word w_i in the sentence chosen at step 3, update their probability

$p_{new}(w_i) = p_{old}(w_i) \cdot p_{old}(w_i)$

Step 5 If the desired summary length has not been reached, go back to Step 2.

SumBasic

Original Features:

- Semantic content units,
- a sentence weighting mechanism based on word frequency
- a mechanism to adjust weights after each selection

Conclusion:

- This method appears more effective than having a separate component which removes duplication.

SumBasic

Our Basic implementations:

- Using word frequency for the calculation of sentence weights.
- Variations: stopwords, bigram, tfidf
  - Suitable mechai

Basic Observations:

- In case of `-no-stopwords` and `prob`, SCUs are usually not among the top-scored sentences
- In case of `-no-stopwords`, dialogue sentences are often selected.
  - Such as s/he said, s/he told me that, ...
- Duplicates

SumBasic, unigram, - with-stopwords, prob

- This method retains common stopwords (e.g., "the," "of," "to") which are often omitted in keyword extraction but can be important for understanding sentence structure and meaning.

- The selected sentences tend to be more narrative or expressive, providing emotional or descriptive insights. This suggests a focus on the overall readability and narrative flow, capturing moments of emotional expression or description.

SumBasic, unigram, -no-stopwords, prob

- By omitting stopwords, this method seemed to be focusing on key content words, potentially highlighting more specific or actionable information.

- The sentences chosen tend to contain direct quotes and actions, emphasizing individuals' reactions and interactions. This method seems to prioritize direct speech and specific statements, possibly reflecting the immediacy or personal perspectives within the document.

SumBasic, unigram, - with-stopwords, tfidf

- Using TF-IDF while including stopwords might balance the focus between common language usage and the importance of specific terms within the context of the larger document set.

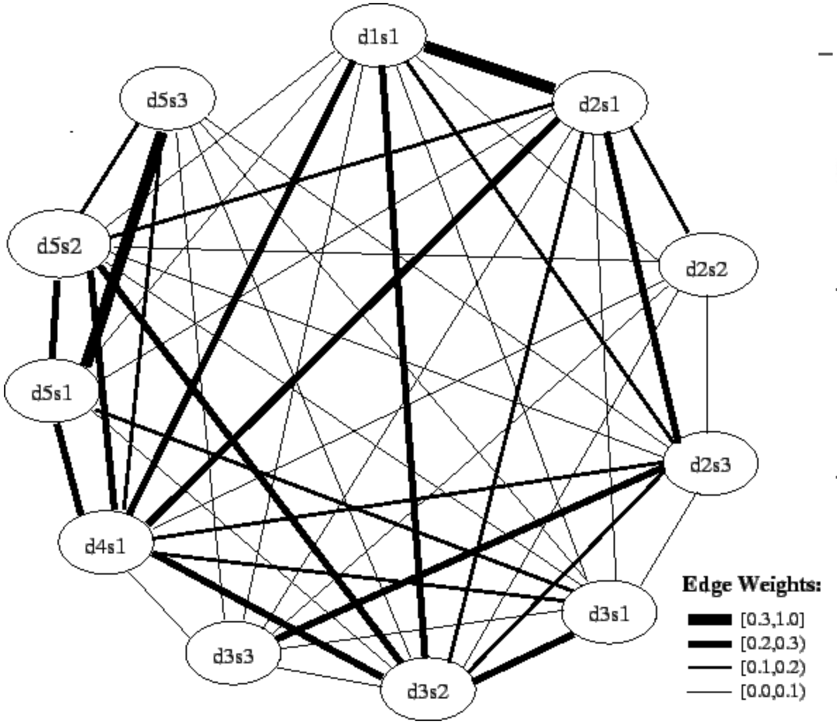
- The selected sentences here concentrate more on the community and location aspects, such as references to Columbine and the communal response. This suggests an emphasis on the thematic significance related to place and communal identity.

SumBasic, unigram, -no-stopwords, tfidf

- This approach focuses on identifying terms that are important within the document but not commonly used across other documents, without the influence of stopwords.

- The sentences selected are more varied, touching on both specific events and broader reactions. The focus is on distinctive information that sets this document apart from others, including specific references to the Columbine incident and its broader implications.

LexRank: A stochastic graph-based method for computing relative importance



- Each sentence is represented by a TF matrix
- Then calculate the cosine similarity by

$$\text{idf-modified-cosine}(x, y) = \frac{\sum_{w \in x,y} \mathrm{tf}_{w,x}\mathrm{tf}_{w,y}(\mathrm{idf}_w)^2}{\sqrt{\sum_{x_i \in x}(\mathrm{tf}_{x_i,x}\mathrm{idf}_{x_i})^2} \times \sqrt{\sum_{y_i \in y}(\mathrm{tf}_{y_i,y}\mathrm{idf}_{y_i})^2}}$$

- Create markov matrix, by zeroing out similarities lower than threshold, and ensure each row of the final matrix sum to 1
- Use power method, to compute the stationary distribution, get final score

LexRank: Examples

```
The nation and the world have joined in grieving for the students of Columbine  Gore said
The teachers are very anxious to see their students again   she said
The community outpouring has touched some Columbine students
The young killers of Columbine High School do not stand for the spirit of America   Gore said
Jefferson County school officials said Columbine s 1800 students would return to classes Thursday a few miles south at Chatfield High School  a school originally
```

```
Through their lawyers  the officers have said they thought Diallo had a gun
The officers in the Diallo case did not testify before the grand jury
When the officers saw afterward that Diallo was unarmed  the people knowledgeable about the case said  at least one of the officers wept
The officers were indicted in March
The defense lawyers called their clients  political prisoners   and said they would show the incident was  a tragedy  not a crime   because the officers had  rea
```

```
The Qinling pandas are believed to have separated from the giant panda about 50000 years ago  Chinese researchers said
Wang said  Shaanxi has so far established 13 giant pandas protection zones and nature reserves focused on pandas  habitats
China has 163 giant pandas in captivity
The bamboo blooming in the early 1980s caused the deaths of about 250 giant pandas
The latest national survey on giant panda  which was organized by the State Forestry Administration  showed that the population and habitat of the pandas in Sichua
```

LexRank: Bad Example

```
This is a serious storm   Suiter said
This is a serious storm   Suiter said
We are going to feel it   he said
It s very scary   Bush said
It s beautiful   said Penalver
```

- It is about the Floyd storm, but it didn’t capture the name of it
- Probably due to similarity threshold being too relaxed

LexRank: Bad Example

```
the 36 workers aboard the platform pumped heavy drilling mud into the leaking injection well .
norway is the world 's third-largest oil exporter after saudi arabia and russia .
norway is the world 's third-largest oil exporter after saudi arabia and russia .
norway is the world 's third-largest oil exporter after saudi arabia and russia .
norway is the third crude exporter after saudi arabia and russia and produces nearly three million barrels per day .
marathon was similarly mum on when production aboard its platform would restart .
```

- Sentence duplication

Evaluation

- Command line format
  - python(3) evaluation.py —model_summary_dir -hypothesis_summary_dir > out_file
- For sumBasic
  - ./rouge.sh ../outputs/sumbasic_unigram_noStop_probability/ ../outputs/sumbasic_unigram_withStop_probability/ sumbasic_unigram_nostop_unigram_withstop_output.txt
  - ./rouge.sh ../outputs/sumbasic_unigram_noStop_probability/ ../outputs/sumbasic_bigram_withStop_probability/ sumbasic_unigram_nostop_bigram_withstop_output.txt
  - **NB:** make sure to include the / at the end of target and prediction directory
- For LLR
  - ./rouge.sh ../outputs/LLR_summaries_stopwords_0.05/ ../outputs/LLR_summaries_stopwords_0.05 llr_stop_nostop_0.05_output.txt

ROUGE results

Method	Document Directories	ROUGE-1 Avg Precision	ROUGE-1 Avg Recall	ROUGE-1 Avg F-measures	ROUGE-2 Avg Precision	ROUGE-2 Avg Recall	ROUGE-2 Avg F-measures	ROUGE-L Avg Precision	ROUGE-L Avg Recall	ROUGE-L Avg F-measures
sumBasic	unigram_withStop	0.238160087	0.2704799348	0.2526288261	0.03551997826	0.04039093478	0.0376996087	0.1377333913	0.1562575	0.146023087
sumBasic	bigram_withStop	0.2356884783	0.2492068478	0.2416698043	0.03615932609	0.03827063043	0.03710241304	0.1378818913	0.1456656739	0.1413234783
sumBasic	unigram_noStop	0.1780128043	0.3489160652	0.2353995652	0.02217080435	0.04347597826	0.02932480435	0.08792171739	0.1723571739	0.1162661739
LLR	stopwords_0.05	0.2425946522	0.196826413	0.2164867609	0.03547223913	0.02893330435	0.03177465217	0.1363083043	0.1104728261	0.1215444783
LLR	nostop_0.05	0.2449904348	0.1956218913	0.2166606739	0.03493241304	0.02782947826	0.03085615217	0.1357332826	0.1082958696	0.1199673261
LexRank	LexRank_summary_length_5	0.2782406304	0.309848913	0.2854538043	0.05004954348	0.0596886087	0.05329119565	0.149613	0.1639114783	0.1520022609