Learning Single-Document Summarization with Compression and Anaphoricity Constraints

Berkeley NLP

Greg Durrett, Taylor Berg-Kirkpatrick, and Dan Klein

{gdurrett,tberg,klein}@cs.berkeley.edu

Introduction

Why is single-document summarization harder than multi-document summarization? First, we don't have redundancy of information as a cue. Second, we don't have as many snippets to choose from when generating our summary. We compensate for these difficulties by learning in an end-to-end way and by giving our model flexibility in generation. Our model is trained on a large corpus (100,000 document-summary pairs) and has a rich feature set so it can learn to select important content. Sentence compression arising from syntactic and discourse parsing lets us pack more content in summaries, and constraints on anaphora ensure that critical segments of text are understandable even if they contain pronouns.

Summarization Process

Find the most relevant document content… (extract and compress sentences)

As the Jewel of the Seas, a luxury cruise ship owned by Royal Caribbean, steamed down the coast toward Bermuda in late October, there were plenty of distractions, including two Olympic-size pools, Latin dance lessons and Boozer Bingo.

[…]

Ms. Jaworski was one of about 150 book lovers aboard the ship for a five-day literary-themed cruise out of Boston. Known as Book It to Bermuda, it is part of a growing genre of cruises called Ship Lit. Often sponsored by publishers, the cruises feature popular authors who give readings and seminars to boatfuls of book lovers.

[…]

Extraction options

…such that it can be rendered grammatically… (each sentence has many possible compressions)

1 *Known as Book It To Bermuda, it is part of a growing genre of cruises called Ship Lit.*

2 *~~Known as Book It To Bermuda~~, it is part of a growing genre of cruises called Ship Lit.*

…

1 *~~Often sponsored by publishers~~, the cruises feature popular authors who give readings and seminars to boatfuls of book lovers.*

…

Summary to render

…with understandable pronouns. (pronouns might need to be rendered as full mentions to be intepretable)

Known as Book It To Bermuda, it is part of a growing genre of cruises called Ship Lit.

*~~It~~ **Book It To Bermuda** is part of a growing genre of cruises called Ship Lit.*

The cruises feature popular authors who give readings and seminars to boatfuls of book lovers.

Final summary

Model

A feature-based linear scoring function evaluates textual units (spans of sentences that serve as minimal units of extraction).

u_i

… | *it is part of a growing genre of cruises* | …

$$\mathrm{score}(u_i) = w^\top f(u_i)$$

$f(u_i) = \begin{cases} \text{Lexical: } \mathbb{I}[\text{first word=}\textit{it}], \mathbb{I}[\text{contains } \textit{growing}], \ldots \\ \text{Position: } \mathbb{I}[\text{paragraph 2}], \mathbb{I}[\text{sentence 10}], \ldots \\ \text{Centrality: } \mathbb{I}[\text{contains word w/count > 10 in doc}]\ldots \end{cases}$

Sentence compression arises from syntactic and RST parses. Nested deletions yield a tree-structured set of constraints, which resemble dependencies.

both needed for grammaticality

Smith | *aged 57* | *is a claims adjuster* | *with Aetna*

optional, can only include if *Smith* is included

optional, can only include if *adjuster* is

For each pronoun in the document, we must either include its antecedent in the summary or rewrite the pronoun unambiguously.

u_1 coreferent u_2

Known as ***Book It To Bermuda*** | ***it*** *is part of…*

To include u_2, either include u_1 or replace ***it***:

it *is part of* ▶ ***Book It To Bermuda*** *is part of*

Only allowed if high coreference confidence

ILP

Find maximal set of units at most *k* words

$$\max_{\mathbf{x}} \sum_i x_i w^\top f(u_i) \text{ s.t. } \sum_i x_i \mathrm{len}(u_i) \le k$$

$x_i = 1$ indicates that u_i is in the summary

which lead to a grammatical summary

s.t. $x_i \le x_j$ if u_i depends on u_j

(x_i can only be included if x_j is)

where pronouns are understandable

s.t. $x_i \le x_j + r_i$ if u_i has a pronoun with antecedent in u_j

$r_i = 1$ indicates pronoun replacement

Learning and Inference

We train with structured SVM on 100,000 New York Times documents paired with *abstractive* summaries. We use the oracle ROUGE-1 summary within model capacity as a surrogate gold.

Subgradient for one example:

$$\nabla \text{objective} = f(\text{oracle ROUGE summ}) - f(\text{loss-augmented summ})$$

computed with modified ILP objective:

requires new unigram type variables and constraints

$$\max \alpha \left(\sum_i x_i w^\top f(u_i) \right) + \beta(\#\text{ unigram types recalled})$$

(set $\alpha = 0,\ \beta = 1$ for oracle, $\alpha = 1,\ \beta = -1$ for loss-aug)

Results

We evaluate on held-out New York Times data. ROUGE-2 measures content relevance; Mechanical Turk lets us evaluate clarity/grammaticality and pronoun coherence.