SIGIR 2022
Madrid

Beyond Opinion Mining: Summarizing Opinions in Customer Reviews

Reinald Kim Amplayo, Arthur Bražinskas, Yoshi Suhara, Xiaolan Wang, Bing Liu

Google Research · grammarly · Megagon Labs · UIC

Schedule

- 11:00-11:40 Part 1: Introduction
- 11:40-12:10 Part 2: Autoencoders
- 12:10-12:40 Part 3: Synthetic Dataset Creation

- 12:40-12:50 QA
- 12:50-13:20 Break

- 13:20-13:50 Part 4: Low-resource learning
- 13:50-14:05 Part 5: Evaluation and resources
- 14:05-14:20 Part 6: Challenges and opportunities

- 14:20-14:30 QA

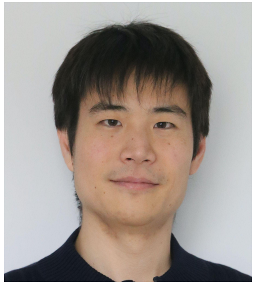
About us

Reinald Kim Amplayo

Google Research

Arthur Bražinskas

Google Research



Yoshi Suhara

grammarly

Xiaolan Wang

Megagon Labs

Bing Liu

UIC U. of Illinois Chicago

Introduction

Motivation: customer reviews

- Users often purchase products **online** (e.g., from Google shopping or Amazon)
- Seek **opinions** of other users expressed in **reviews**
- Use this information for **informed purchasing decisions**

Customer reviews: example

Google Pixel 4 - Just Black - 64GB - Unlocked

Visit the Google Store

906 ratings | 139 answered questions

Climate Pledge Friendly

Price: $373.98

$86.28 Shipping & Import Fees Deposit to United Kingdom Details

Sales taxes may apply at checkout

Available at a lower price from other sellers that may not offer free Prime shipping.

Enhance your purchase

Trade-in and save

Get up to $511.00 added to your Amazon.com Gift Card balance when you trade in your old device(s)

Start your trade-in

Model Name	Pixel 4
Wireless Carrier	Unlocked
Brand	Google
Form Factor	Smartphone
Memory Storage Capacity	64 GB
Operating System	Android

See more

$373.98

$86.28 Shipping & Import Fees Deposit to United Kingdom

Details

Sales taxes may apply at checkout

Delivery **Friday, February 4**

Or fastest delivery **Thursday, January 27**. Order within 7 hrs 5 mins

Deliver to United Kingdom

In Stock.

Qty: 1

Add to Cart

Buy Now

Secure transaction

Ships from Amazon

Sold by Majestueux

Return policy: Eligible for Return, Refund or Replacement

Support: Free Amazon product support included

Add an Accessory:

Amazon Basics $6.82

Customer reviews: example

Ricochet Top Contributor: Photography

★★★☆☆ **Pixel4 - On The Fence With This One - Google Has Cut One too Many Corners Here**

Reviewed in the United States on October 24, 2019

Verified Purchase

Have been using Android over 10 years and carrying Pixels the last 4 years. Love stock Android. I don't usually buy phones at launch, but I caved this year. Don't care about the camera, but if this is important for you, look no further. Like it's predecessors, this phone takes nice photos. Cool design, premium feel, and a modestly updated feature set. Android is still Android after all. The phone does feel good in your hand. Screen is very nice and the 90hz refresh is noticeably smooth. If you've been reading reviews, you've probably read all this elsewhere.

I do want to like this phone and am going to give it a chance, but I'm going to be merciless if the device doesn't stand up to reasonable everyday use. What do I mean by this? We bought two phones but I've only opened one of them to evaluate. Set up was easy. Connect your devices and less than 10 min later, everything was transferred. Apps installed, then they updated. Ready for testing. WiFi and Bluetooth connection were fast and solid.

Customer reviews: challenges

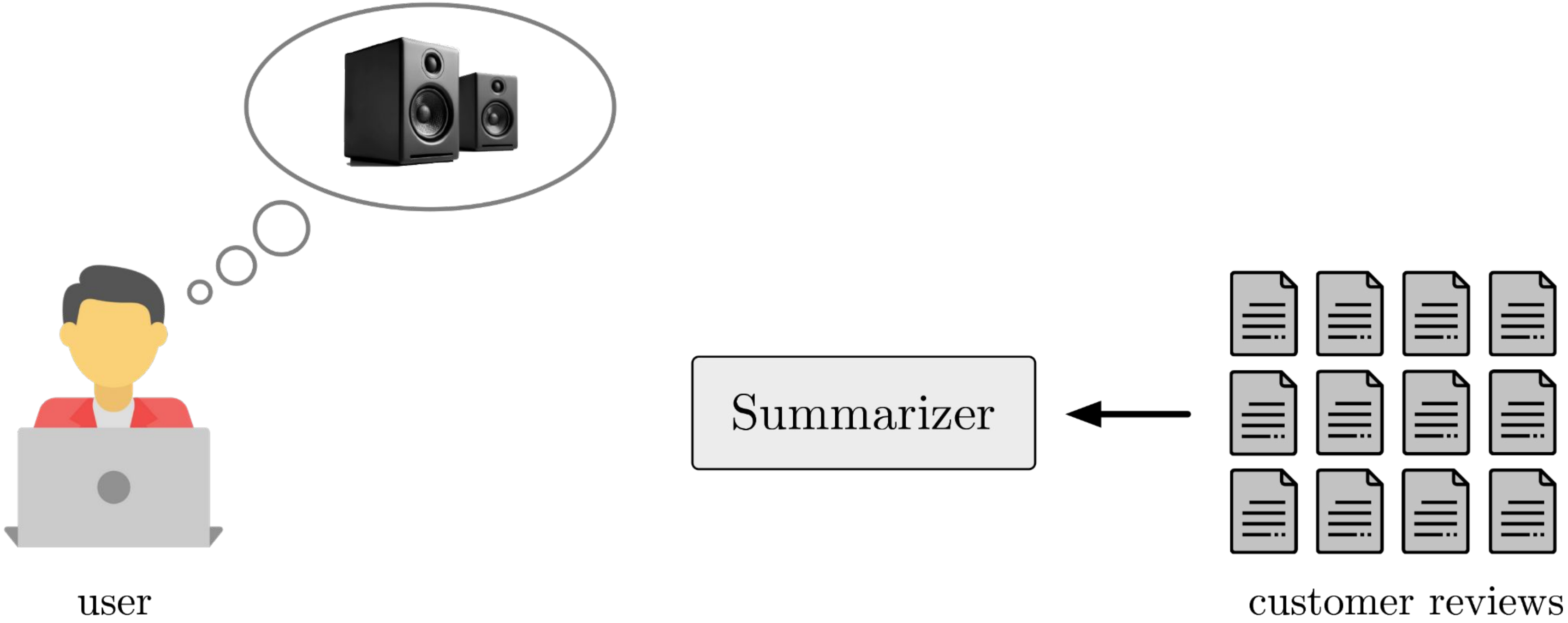
- **Volume:** some products/services have thousands of reviews
- **Absence of explicit structure**: challenging to retrieve reviews discussing particular information
- **Repeated content**: many reviews expressing the same opinions.
- **Uninformative content:** reviews often mention personal experiences not useful for decision making

Automatic summarization

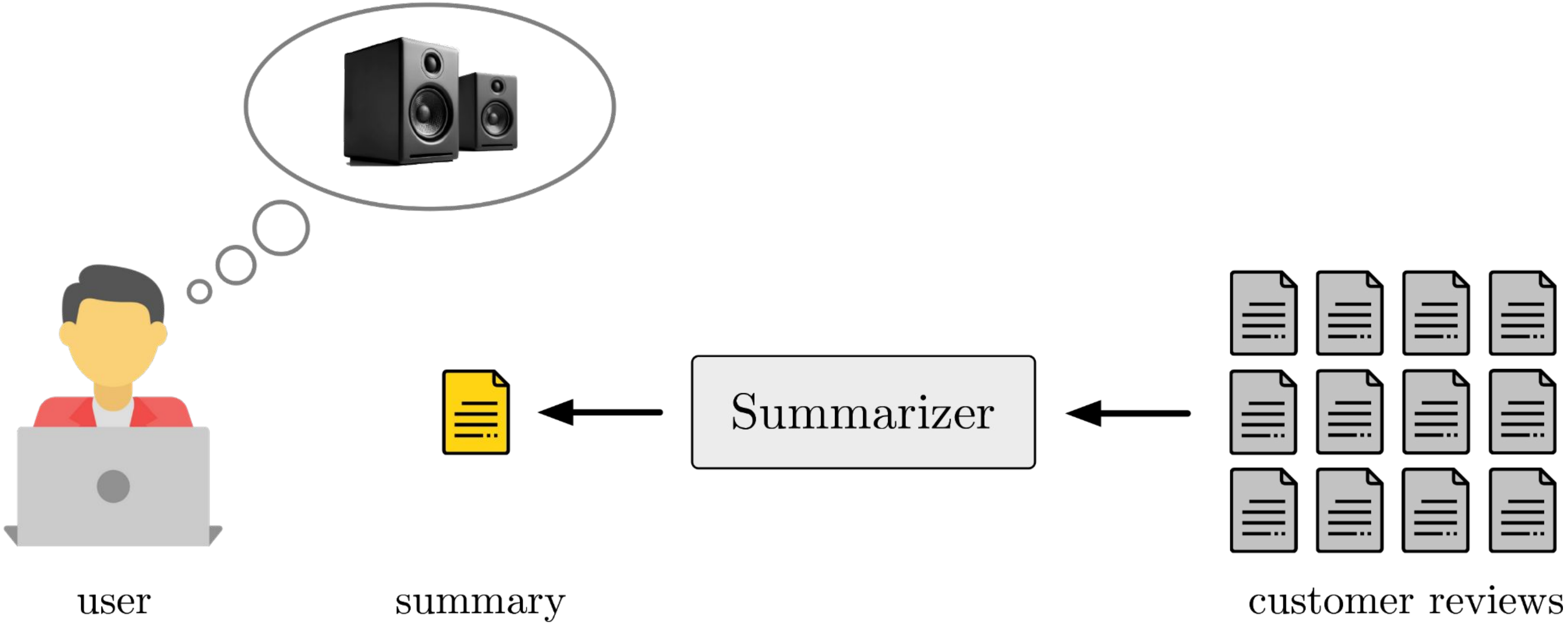
- Automatic summarization can **compress** and **fuse** opinions to short texts
- Summaries are useful for **fast and informed decision making**

Automatic summarization

Automatic summarization



Automatic summarization



Presentation of Opinion Summaries

- Summaries can be presented on web platforms, such as amazon.com
- Vocalized by **voice-based assistants** (e.g., Alexa and Google Assistant)

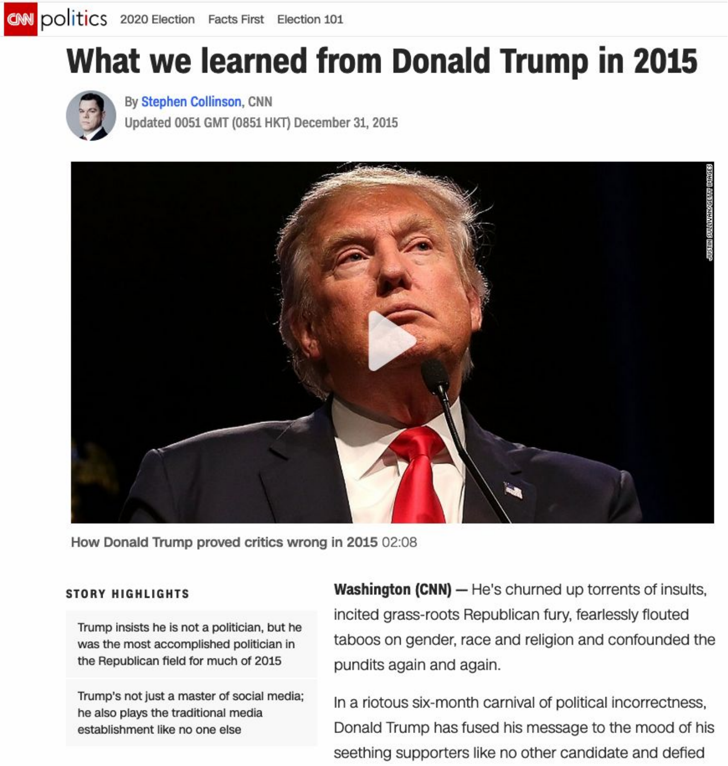
Presentation of Opinion Summaries

- Summaries can be presented on web platforms, such as amazon.com
- Vocalized by **voice-based assistants** (e.g., Alexa and Google Assistant)

News summarization

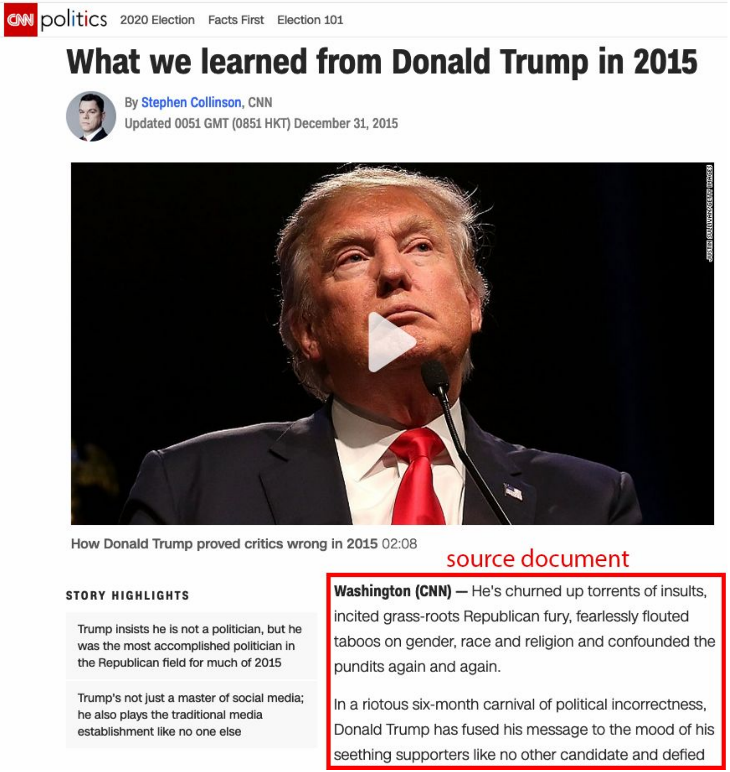
News summarization

- A related branch – **news summarization**
- Mostly summarization of a **single document**



News summarization

- Journalist-written article is used as input/source to the model
- Often more than 800 words long



CNN politics 2020 Election Facts First Election 101

What we learned from Donald Trump in 2015

By Stephen Collinson, CNN

Updated 0051 GMT (0851 HKT) December 31, 2015

How Donald Trump proved critics wrong in 2015 02:08

STORY HIGHLIGHTS

Trump insists he is not a politician, but he was the most accomplished politician in the Republican field for much of 2015

Trump's not just a master of social media; he also plays the traditional media establishment like no one else

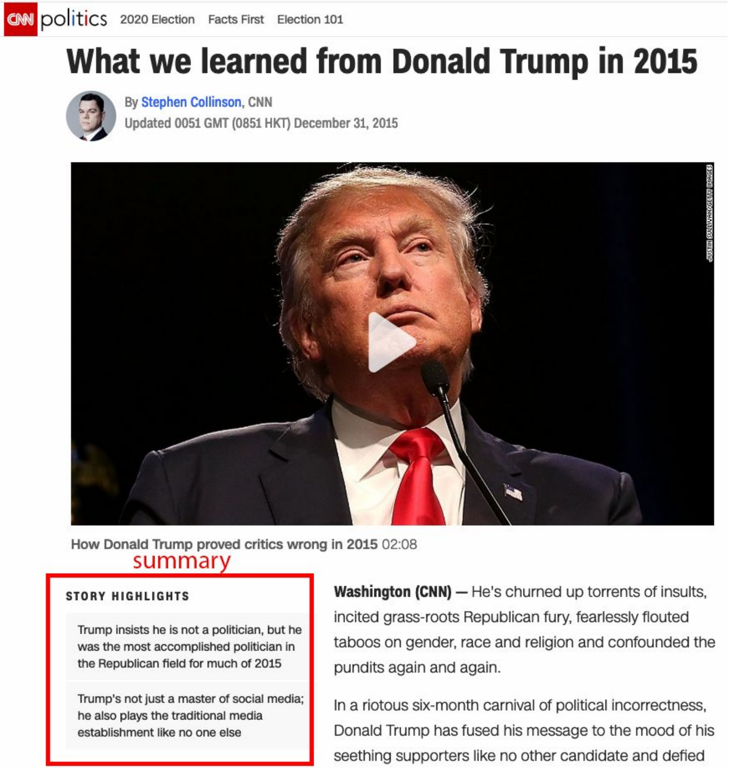
source document

Washington (CNN) — He's churned up torrents of insults, incited grass-roots Republican fury, fearlessly flouted taboos on gender, race and religion and confounded the pundits again and again.

In a riotous six-month carnival of political incorrectness, Donald Trump has fused his message to the mood of his seething supporters like no other candidate and defied

News summarization

- Summaries often are:
  - Story highlights
  - First sentences in the article



News vs opinion summarization

	News	Opinion
Setup	Single-document	Multi-document

News vs opinion summarization

	News	Opinion
Setup	Single-document	Multi-document
Annotated data	1M+ (Grusky et. al. 2018)	100 (Chu and Liu, 2019)

News vs opinion summarization

	News	Opinion
Setup	Single-document	Multi-document
Annotated data	1M+ (Grusky et. al. 2018)	100 (Chu and Liu, 2019)
Data type	Objective facts	Subjective opinions

News vs opinion summarization

	News	Opinion
Setup	Single-document	Multi-document
Annotated data	1M+ (Grusky et. al. 2018)	100 (Chu and Liu, 2019)
Data type	Objective facts	Subjective opinions
First abstractive model	(Nallapati et. al 2016)	(Chu and Liu, 2019)

Extractive and abstractive summarizers

Extractive and abstractive summarizers

- There are **two types** of summarizers
- Extractive: **select** input document (s) sentences to form a summary
- Abstractive: **generate** summaries word-by-word

Extractive summarizers

- Select **salient sentences** in reviews
- Concatenate the **selected sentences** to form a **summary**
- Can be **incoherent** and contain **unimportant details**

Example

DAGOSTINO'S

Italian

Example

Review 1
The stake was cold, and the bread was sour. The server forgot about our order.

Example

Review 1	Review 2
The stake was cold, and the bread was sour. The server forgot about our order.	The waitress was very rude. The pasta was too dry, would not recommend it.

Example

Review 1	Review 2
The stake was cold, and the bread was sour. The server forgot about our order.	The waitress was very rude. The pasta was too dry, would not recommend it.

Example

Review 1	Review 2
The stake was cold, and the bread was sour. The server forgot about our order.	The waitress was very rude. The pasta was too dry, would not recommend it.

Two **aspects** are discussed: food and service

Example

Review 1	Review 2
The stake was cold, and the bread was sour. The server forgot about our order.	The waitress was very rude. The pasta was too dry, would not recommend it.

Extractive summary?

Example

Review 1	Review 2
The stake was cold, and the bread was sour. The server forgot about our order.	The waitress was very rude. The pasta was too dry, would not recommend it.

Extractive summary: The server forgot about our order. The pasta was too dry, would not recommend it.

Example

Review 1	Review 2
The stake was cold, and the bread was sour. The server forgot about our order.	The waitress was very rude. The pasta was too dry, would not recommend it.

Abstractive summary: This restaurant has poor service and food. It is not recommended.

Extractive methods

- Well studied across different summarization branches
- Usually framed as a **tagging problem**:
    - Given a document (s)
    - Select K **summarizing fragments** (e.g., sentences)
    - Concatenate to form a summary

Extractive methods

- The central challenge is how to **represent sentences**
- Need **`rich’ semantic representations** for accurate binary classification

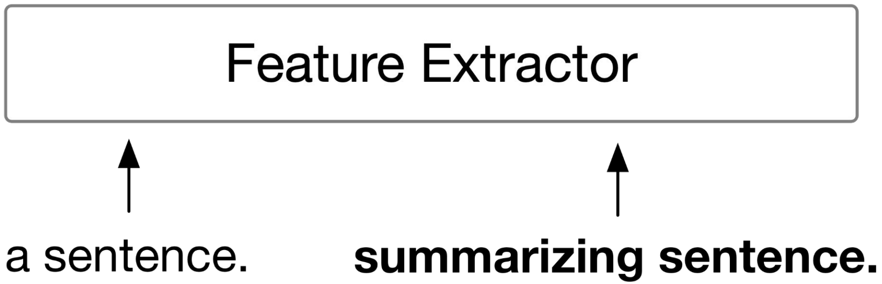
Extractive methods: structure

- The feature extractor is usually a **neural encoder** that produces **sentence semantic representations**
- Such as a Transformer (Vaswani et al., 2017)
- Often pre-trained (Liu and Lapata, 2019)
- The representations are used for **binary classification**

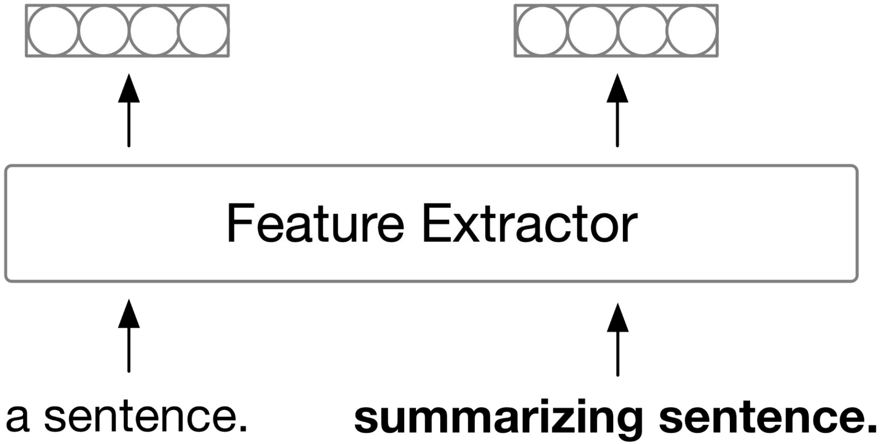
Extractive methods: workflow

a sentence. **summarizing sentence.**

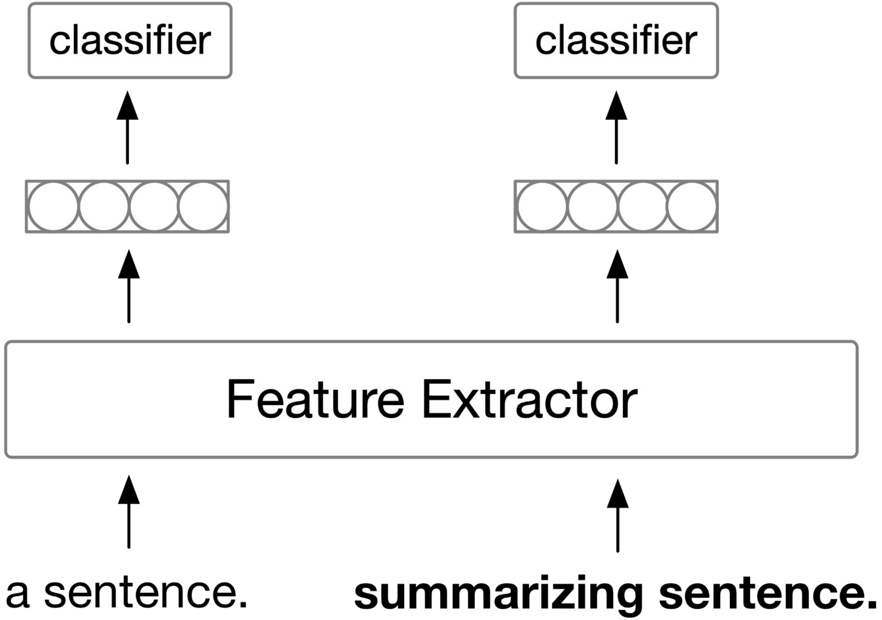
Extractive methods: workflow



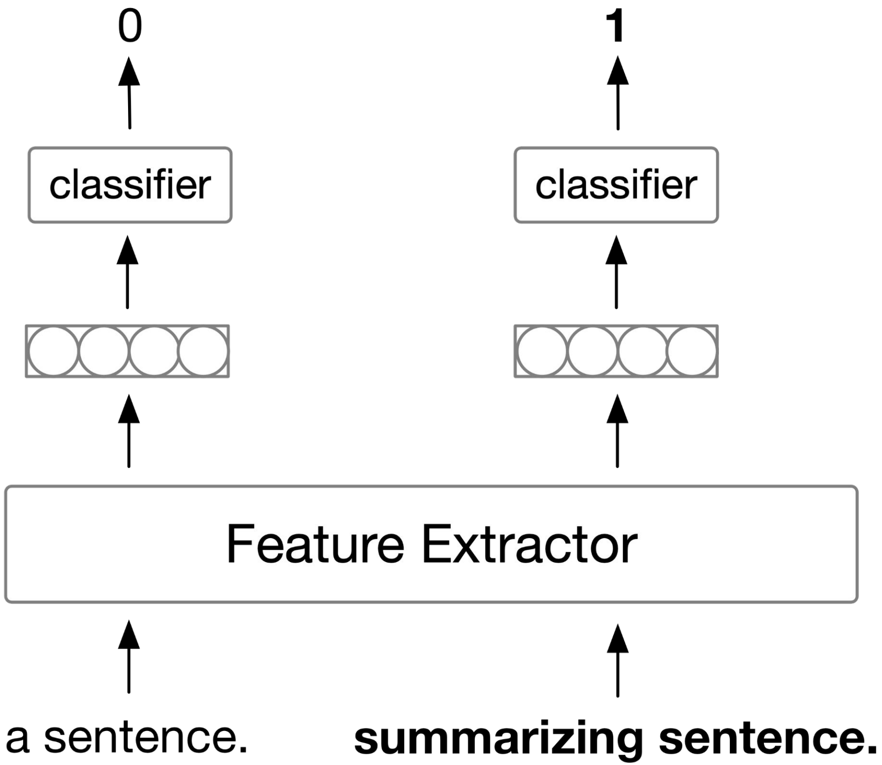
Extractive methods: workflow



Extractive methods: workflow



Extractive methods: workflow



Extractive methods: pros

- `Easy-to-build' models
- Always factually correct summaries
- Faster training and inference
- Less data demanding (often unsupervised or weakly supervised)

Extractive methods: cons

- Incoherent output
- `Jammed' unimportant details
- Inability to abstract information
- Limited vocabulary of words

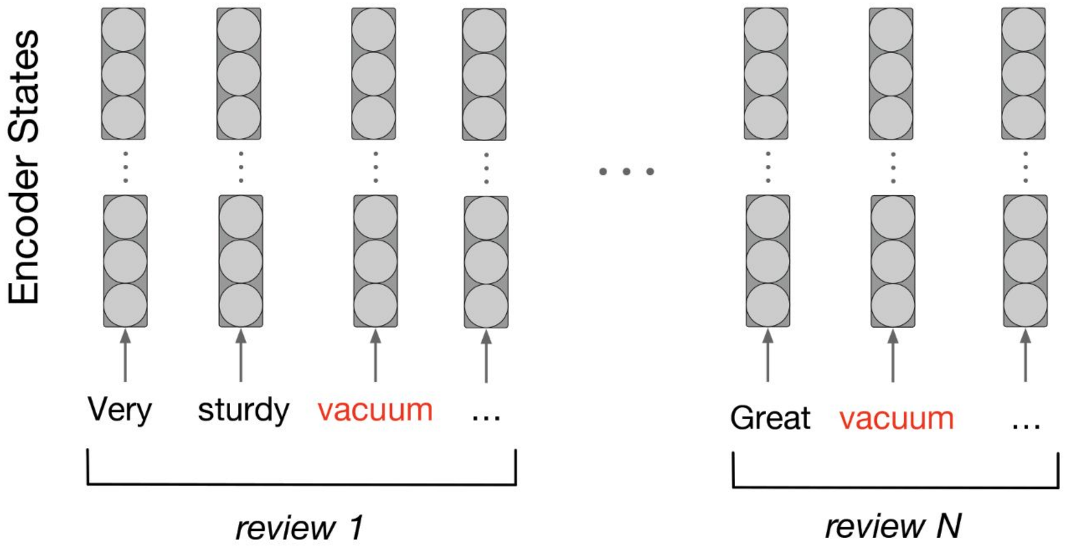
Extractive methods: cons

- Extractive methods are **effective** for news (Nallapati et al., 2016)
- But less effective for **controversial** corpora (Carenini and Cheung, 2006)
  - Customers often **disagree** about product aspects/details
  - There are often no **`summarizing’** sentences

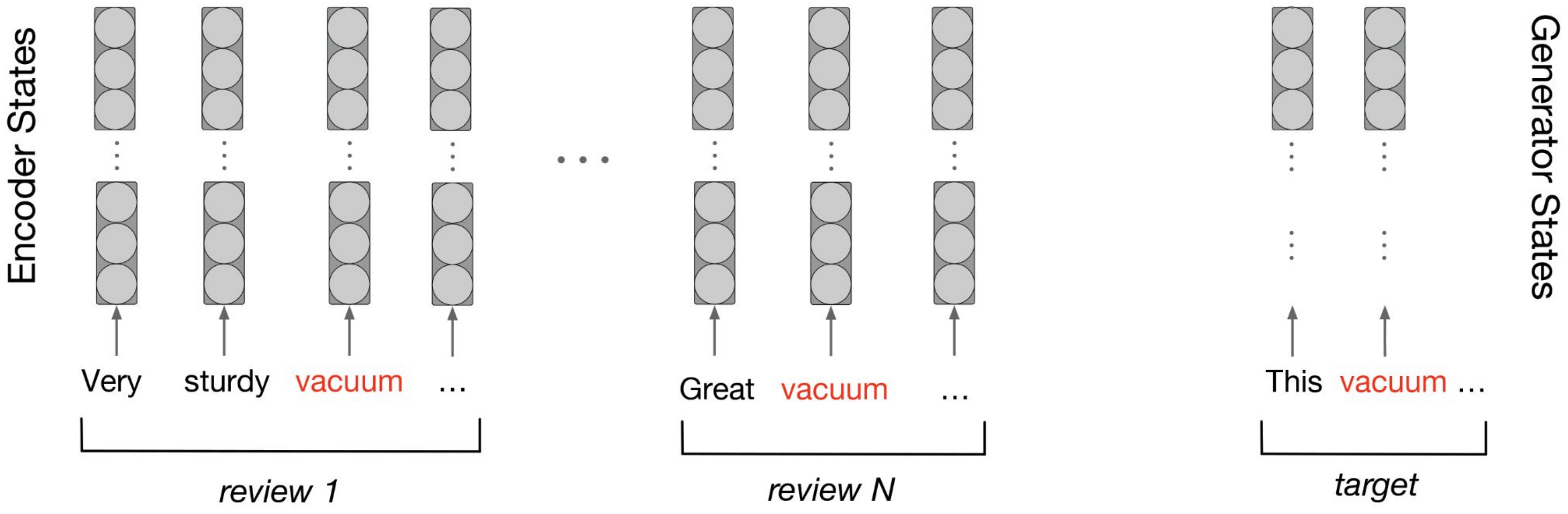
Abstractive methods

- Based on the **encoder-decoder** architecture
- Generate text word-by-word (Nallapati et al., 2016, Paulus et al. 2017)
- Can **compress** and **fuse** (Lebanoff et al., 2019)
- Can deal with **conflicting information** in different user reviews

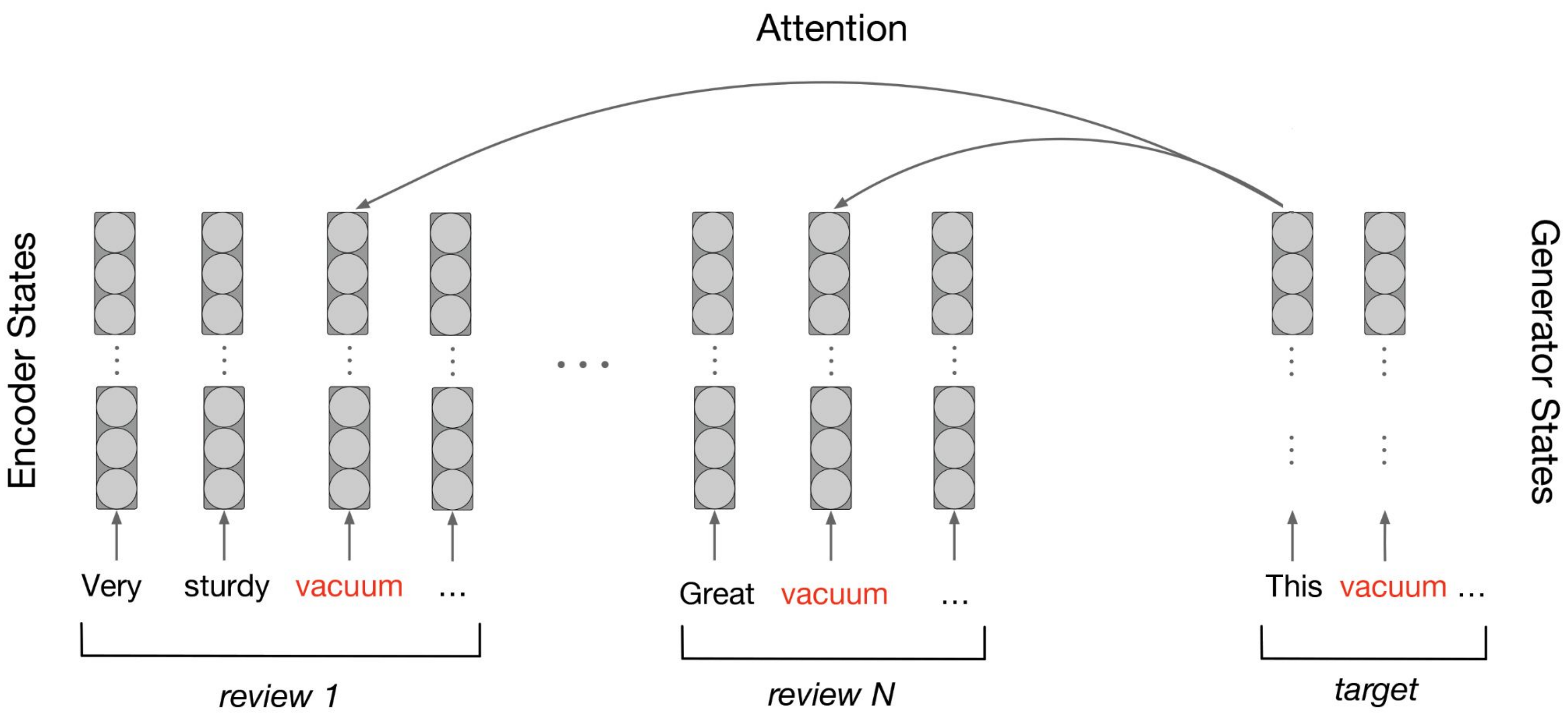
Abstractive methods: workflow



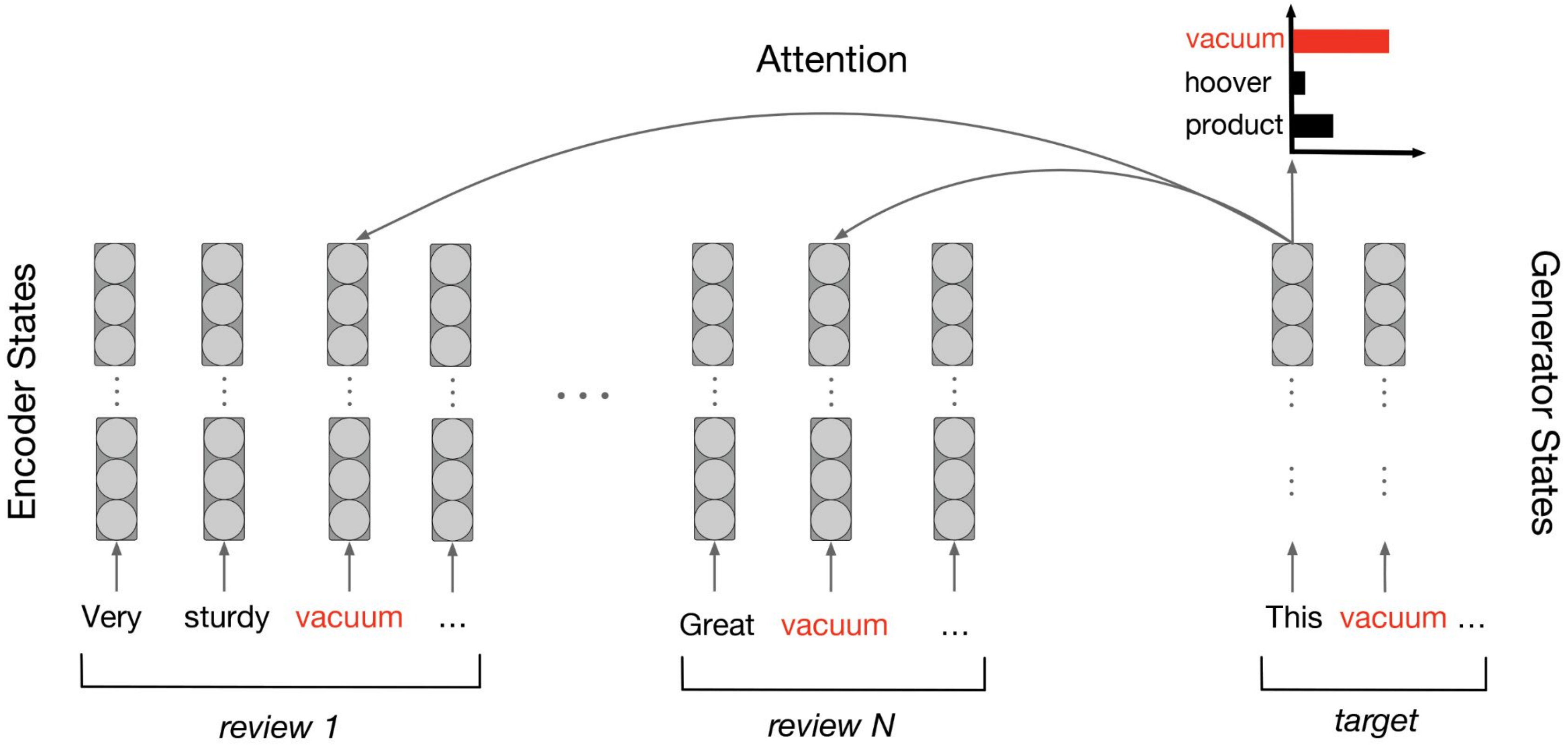
Abstractive methods: workflow



Abstractive methods: workflow



Abstractive methods: workflow



Abstractive summarizers: challenges

- More data demanding
- Often more memory and computationally demanding
- Have a number of open problems:
  - How to encode many documents using deep encoders
  - How to generate input faithful summaries (Maynez et al., 2020)
  - …

History of Opinion Summarization

1. Kim, Hyun Duk, Kavita Ganesan, Parikshit Sondhi, and ChengXiang Zhai. "Comprehensive review of opinion summarization." (2011).

Structured Opinion Summaries

Aspect-Sentiment Table Summaries

- Create a table showing overall sentiment scores for each aspect of an entity
- Information is **easier to digest** as reading is not required
- **Too much abstraction** from reviews, giving **no explanations** for the predicted sentiments

Movie	Screenplay	History	Feeling	Direction	Patriotism	Character	Acting
The Admiral: Roaring Currents	86.45	88.21	92.42	**73.25**	95.06	77.83	94.93
Kundo: Age of the Rampant	66.02	71.03	68.71	58.75	—	58.61	80.41
The Attorney	—	80.46	89.37	–	92.97	73.96	**95.16**
The Suspect	71.53	—	82.34	62.77	—	65.08	94.59
A Hard Day	**91.03**	—	93.92	—	—	90.47	98.47
The Fatal Encounter	62.99	76.69	70.40	49.05	84.11	—	85.01

1. Wu, Haibing, Yiwei Gu, Shangdi Sun, and Xiaodong Gu. "Aspect-based opinion summarization with convolutional neural networks." In *IJCNN*, pp. 3157-3163. IEEE, 2016.
2. Amplayo, Reinald Kim, and Min Song. "An adaptable fine-grained sentiment analysis for summarization of multiple short online reviews." *Data & Knowledge Engineering* 110 (2017): 54-67.

Structured Opinion Summaries

Statistical or Feature-based Summaries

- Preprocess reviews using aspect/feature identification and sentiment classification models
- Present results as a cluster of opinions
- Opinions are **organized better** than just showing a list of reviews
- Users still would **need to read *a lot of* text** to answer the question “why positive?”

Feature1: **picture**
Positive: 12

- The pictures coming out of this camera are amazing.
- Overall this is a good camera with a really good picture clarity.

…
Negative: 2

- The pictures come out hazy if your hands shake even for a moment during the entire process of taking a picture.
- Focusing on a display rack about 20 feet away in a brightly lit room during day time, pictures produced by this camera were blurry and in a shade of orange.

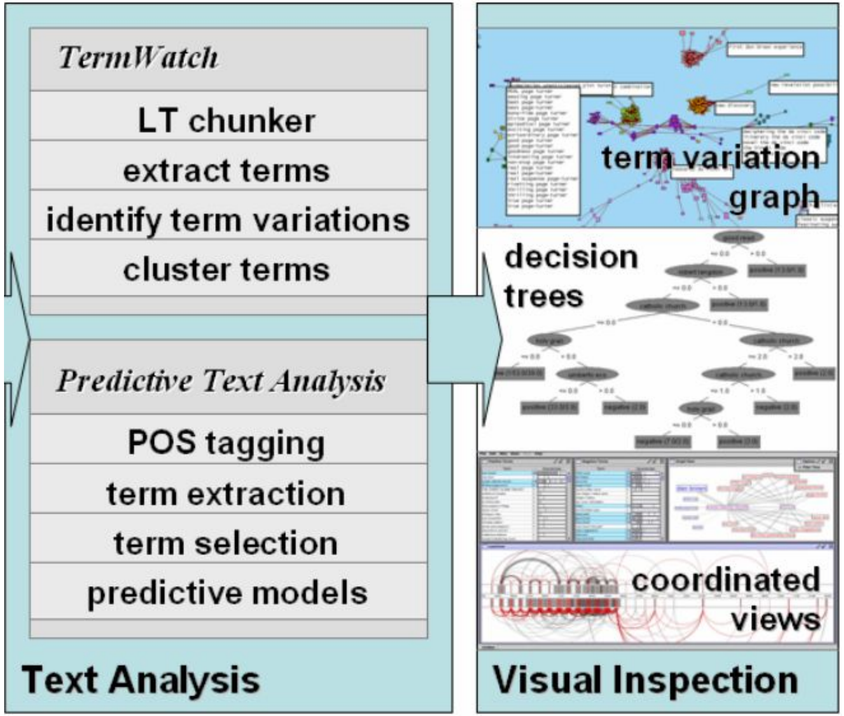
Feature2: **battery life**
…

1. Hu, Minqing, and Bing Liu. "Mining and summarizing customer reviews." In *KDD*, pp. 168-177. 2004.
2. Hu, Minqing, and Bing Liu. "Opinion extraction and summarization on the web." In *AAAI*, vol. 7, pp. 1621-1624. 2006.
3. Zhuang, Li, Feng Jing, and Xiao-Yan Zhu. "Movie review mining and summarization." In *CIKM*, pp. 43-50. 2006.

Structured Opinion Summaries

Visual Summaries

- Create graph structures from reviews to visualize information
- **Useful for businesses** to analyze trends and trajectories of their products
- May **not be very friendly to end users**



1. Chen, Chaomei, Fidelia Ibekwe-SanJuan, Eric SanJuan, and Chris Weaver. "Visual analysis of conflicting opinions." In 2006 IEEE Symposium On Visual Analytics Science And Technology, pp. 59-66. IEEE, 2006.

Structured vs. Textual Opinion Summaries

Structured Summaries

- Commonly used as data visualization tools to analyze reviews
- May either contain insufficient or overloaded information

Textual Summaries

- Easily understandable outputs that can be presented to end **users**
- Brief, comprehensible, and easily digestible by **users**

1. Murray, Gabriel, Enamul Hoque, and Giuseppe Carenini. "Opinion summarization and visualization." In Sentiment Analysis in Social Networks, pp. 171-187. Morgan Kaufmann, 2017.
2. Moussa, Mohammed Elsaid, Ensaf Hussein Mohamed, and Mohamed Hassan Haggag. "A survey on opinion summarization techniques for social media." Future Computing and Informatics Journal 3, no. 1 (2018): 82-109.

Pre-Neural Opinion Summarization Approaches

Three types of approaches

- **NLG-based approaches** use a traditional NLG pipeline to select and plan content using *content planning* and produce summaries with *surface realization*
  - SEA[1]
- **Graph-based approaches** build a graph (can be sentence- or token-level) from reviews and use it to infer salient content from reviews
  - LexRank[2]
  - Opinosis[3]
- **Topic model based approaches** infer aspect/sentiment distributions from a corpus reviews using topic models, and extract opinions using them
  - Topic-Sentiment Mixture model (TSM)[4]
  - Multi-Aspect Sentiment model (MAS)[5]

1. Carenini, Giuseppe, Raymond Ng, and Adam Pauls. "Multi-Document Summarization of Evaluative Text." In *EACL*, pp. 305-312. 2006.
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. Ganesan, Kavita, ChengXiang Zhai, and Jiawei Han. "Opinosis: A Graph Based Approach to Abstractive Summarization of Highly Redundant Opinions." In *COLING*, pp. 340-348. 2010.
4. Mei, Qiaozhu, Xu Ling, Matthew Wondra, Hang Su, and ChengXiang Zhai. "Topic sentiment mixture: modeling facets and opinions in weblogs." In *WWW*, pp. 171-180. 2007.
5. Titov, Ivan, and Ryan McDonald. "A joint model of text and aspect ratings for sentiment summarization." In *ACL*, pp. 308-316. 2008.

NLG-based Approaches

Summarizer of Evaluative Arguments (SEA)[1]:
> Transforms a feature-based summary into a textual summary using an traditional natural language generation pipeline: content selection and surface realization

Feature1: **picture**
Positive: 12
- The pictures coming out of this camera are amazing.
- Overall this is a good camera with a really good picture clarity.

…
Negative: 2
- The pictures come out hazy if your hands shake even for a moment during the entire process of taking a picture.
- Focusing on a display rack about 20 feet away in a brightly lit room during day time, pictures produced by this camera were blurry and in a shade of orange.

Feature2: **battery life**
…

Feature-based Summary

overall reception:
exceedingly well
superior
customer types:
novice photographers
semi-professional
mentioned attributes:
picture
battery life
…

Selected Content

The Canon G3 was received **exceedingly well**. Consumer reviews from **novice photographers** to **semi-professional** all listed **an impressive** number of attributes, they claim makes this camera **superior** in the market. …

Content-Filled NLG Template

1. Carenini, Giuseppe, Raymond Ng, and Adam Pauls. "Multi-Document Summarization of Evaluative Text." In *EACL*, pp. 305-312. 2006.

Graph-based Approaches

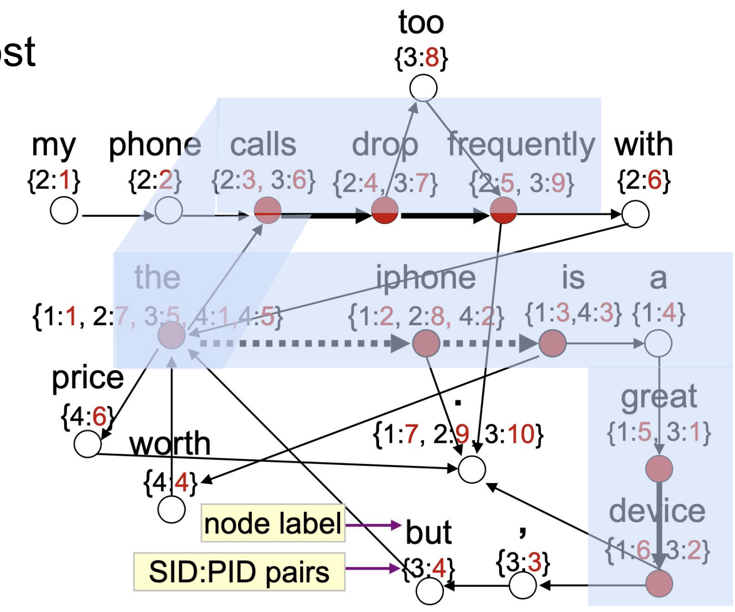
Opinosis[1]:
> Creates a graph from tokens and traverses the most frequent edges to produce a summary.

<u>Reviews</u>

1. The iPhone is a great device.
2. My phone calls drop frequently with the iPhone.
3. Great device, but the calls drop too frequently.
4. The iPhone is worth the price.

<u>Summary</u>

the iphone is a great device.
the calls drop frequently.



Token Neighborhood Graph

1. Ganesan, Kavita, ChengXiang Zhai, and Jiawei Han. "Opinosis: A Graph Based Approach to Abstractive Summarization of Highly Redundant Opinions." In *COLING*, pp. 340-348. 2010.

Topic Model based Approach

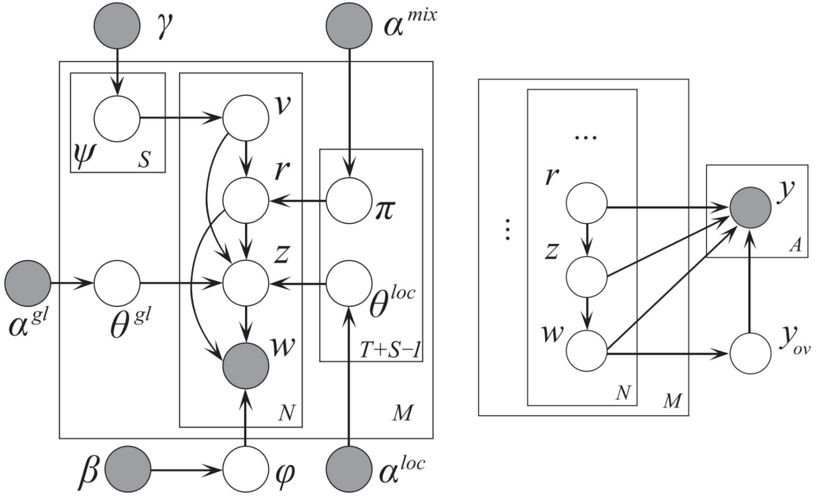
Multi-Aspect Sentiment (MAS)[1]:
> Extends LDA topic model to have local/global topics and predict aspect-specific scores
> Uses the model to sample aspect-specific sentences

w - word
z - topic
r - global/local switch
v - window of text
y - aspect rating

<u>Summary</u>

Nikos' Fine Dining

Food	4/5	"Best fish in the city", "Excellent appetizers"
Decor	3/5	"Cozy with an old world feel", "Too dark"
Service	1/5	"Our waitress was rude", "Awful service"
Value	5/5	"Good Greek food for the $", "Great price!"



1. Titov, Ivan, and Ryan McDonald. "A joint model of text and aspect ratings for sentiment summarization." In *ACL*, pp. 308-316. 2008.

Modern Opinion Summarization

- The modern opinion summarization is based on **neural networks**
  - Better representation learning via auto-encoding (self-supervised objectives)
  - Models have higher capacity (more parameters)
  - Can efficiently learn from large amounts of data (customer reviews) – training on GPUs

Modern Opinion Summarization: Challenges

- The lack of **annotated data**
  - Expensive annotation
  - Hard to annotate prevalent opinions from massive number of reviews
  - Difficult to write fluent and coherent summaries
- Modelling many reviews
  - Some products can have thousands of reviews

Roadmap

Opinion Summarization Approaches

- Autoencoders
- Synthetic dataset creation
- Low-resource learning

Evaluation and Resources

- Metrics
- Datasets

Challenges and Opportunities