Opinosis:
A Graph Based Approach to Abstractive Summarization of Highly Redundant Opinions

Kavita Ganesan, ChengXiang Zhai, Jiawei Han
University of Illinois @ Urbana Champaign
Overall aspect based ratings for entity [Lu et al., 2009; Lerman et al., 2009;..]

Customer Reviews

Average Customer Rating

- **Appearance**: ★★★★★ (1,213)
- **Ease of use**: ★★★★★ (1,212)
- **Portability**: ★★★★★ (1,202)
- **Sound quality**: ★★★★★ (1,196)

Most Helpful Customer Reviews

3,677 of 3,770 people found the following review helpful.

★☆☆☆☆ **WARNING** for new 8GB 3G owners and those on a budget...

By **Hassan B. Bn Hadhram** - See all my reviews

**Amazon Verified Purchase** *(What's this?)*

This review is from: Apple iPod touch 8 GB (2nd Generation)

Before I start let me just tell you "what's New" with the iPod touch Third generation :

- Faster Cpu/Double the ram/Better graphic (faster Boot time/faster loading is all what I did notice)
- Double the storage for the same old price
- Voice control (I'll explain it in a second)
- Latest firmware for free
- New Earbuds with built in remote+Microphone (So you can use voice control)

And that is everything~ depends on your needs upgrading from 2G to 3G might be not worth it.

*Important Note*: only the New iPod touch 32GB/64GB are third generation ~ (8GB is repacked 2G) Details below.
Opinion Summarization Today...

Customer Reviews

Average Customer Rating

<table>
<thead>
<tr>
<th>Rating</th>
<th>Number of Reviews</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 star</td>
<td>1,040</td>
</tr>
<tr>
<td>4 star</td>
<td>222</td>
</tr>
<tr>
<td>3 star</td>
<td>63</td>
</tr>
<tr>
<td>2 star</td>
<td>9</td>
</tr>
<tr>
<td>1 star</td>
<td>0</td>
</tr>
</tbody>
</table>

Appearance .................................. ★★★★★ (1,213)
Ease of use .................................. ★★★★★ (1,212)
Portability .................................. ★★★★★ (1,202)
Sound quality .................................. ★★★★★ (1,196)

To know more: read many sentences

Structured format ➔ useful, but not enough!
We need....
supporting textual summary!
Properties of unstructured opinion summaries

- Summarize the major opinions
  - What are the major complaints/praise in an aspect?

- Concise
  - Easily digestible
  - Viewable on smaller screens

- Readable
  - Easily understood
How to generate such summaries?
Extractive Summarization

- Widely studied for years
  [Radev et al.2000; Erkan & Radev, 2004; Mihalcea & Tarau, 2004...]

- But, not suitable for:
  1. generating concise summaries
  2. summarizing highly redundant text
Problem 1: Bias with limit on summary size

Ex: Opinion sentences about iPhone’s Battery

1. The iPhone’s battery lasts long...have to charge it once every few days.

2. iPhone’s battery is bulky but it is cheap..

3. iPhone’s battery is bulky but it lasts long!
Ex: Opinion sentences about iPhone’s Battery

1. The iPhone’s battery lasts long...have to charge it once every few days.

Tendency of missing out information
Extractive summarization

Problem 2: Verbose
- May contain irrelevant information
- Not suited for smaller devices
An Ideal Summary

“**The iPhone’s battery lasts long and is cheap** but its **bulky.**

- Important information summarized
- Concise
- Readable
An Ideal Summary

“The iPhone’s battery lasts long and is cheap but its bulky.

Extractive ✗
Abstractive ✓
Abstractive Summarization

- HARD!!

- Some methods require manual effort
  [DeJong1982] [Radev and McKeown1998] [Finley and Harabagiu2002]
  - Need to define \texttt{templates} to be filled

- Some methods rely on deep NL understanding
  [Saggion and Lapalme2002] [Jing and McKeown2000]
  - Domain dependent
  - Impractical – high computational costs
Our Method: Opinosis

- ‘Shallow’ abstractive summarizer

- Generates **concise summaries** using:
  - existing text
  - inherent redundancies

- Uses **minimal** external knowledge
  - lightweight
Opinosis: High Level Overview
Opinosis: High Level Overview

Set of sentences:

• **Topic** specific (ex. battery life of ipod)
• **POS** annotated
Opinosis: High Level Overview

Set of sentences:
- **Topic specific** (ex. battery life of ipod)
- **POS annotated**

**Input**

**Step 1:** Generate **graph representation of sentences (Opinosis-Graph)**

- My phone calls drop frequently with the iphone is a great device.
Opinosis: High Level Overview

Set of sentences:
- **Topic specific** (ex. battery life of ipod)
- **POS annotated**

**Input**

Step 1: Generate graph representation of sentences (**Opinosis-Graph**)

Step 2: Find **promising paths** (candidate summaries) & score these candidates
Opinosis: High Level Overview

Set of sentences:
- **Topic** specific (ex. battery life of iPod)
- **POS** annotated

Input

The iPhone is a great device, but calls drop frequently.

Step 3: Select top scoring candidates as final summary

Step 1: Generate graph representation of sentences (Opinosis-Graph)

Step 2: Find promising paths (candidate summaries) & score these candidates
Step 1: Building the Opinosis-Graph
Assume:

- 2 sentences about “call quality of iphone”
  1. My phone calls drop frequently with the iPhone.
  2. Great device, but the calls drop too frequently.

Opinosis-Graph is empty
1. My phone calls drop frequently with the iPhone.
1. *My phone calls drop frequently with the iPhone.*
1. My phone calls drop frequently with the iPhone.
1. *My phone calls drop frequently with the iPhone.*
1. My phone calls drop frequently with the iPhone.
1. My phone calls drop frequently with the iPhone.

Building Opinosis-Graph
1. *My phone calls drop frequently with the iPhone.*
1. My phone calls drop frequently with the iPhone.
1. *My phone calls drop frequently with the iPhone.*
1. My phone calls drop frequently with the iPhone.
1. My phone calls drop frequently with the iPhone.
1. My phone calls drop frequently with the iPhone.
2. Great device, but the calls drop too frequently.
2. Great device, but the calls drop too frequently.

my phone calls drop frequently with the iphone

great device , but
2. Great device, but the calls drop too frequently.
2. Great device, but the calls drop too frequently.
2. Great device, but the calls drop too frequently.
2. Great device, but the calls drop too frequently.
2. Great device, but the calls drop too frequently.
2. Great device, but the calls drop too frequently.
2. Great device, but the calls drop too frequently.
Building Opinosis-Graph

Graph is now ready for Step 2!
3 Important Properties of the Opinosis-Graph
Property 1

Naturally captures redundancies

my phone calls drop frequently with the iphone
great device , but
too
1:1 1:2 1:3, 2:6 1:4, 2:7 1:5, 2:9 1:6 1:7, 2:5 1:8 1:9, 2:10 2:1 2:2 2:3 2:4
Property 1

Naturally captures redundancies

Path shared by 2 sentences naturally captured by nodes
Property 1

Naturally captures redundancies

calls drop frequently

1:3, 2:6
1:4, 2:7
1:5, 2:9

Naturally captures redundancies

Easily discover redundancies for high confidence summaries
Property 2

Captures gapped subsequences

1. My phone *calls drop frequently* with the iPhone.
2. Great device, but the *calls drop too frequently*. 
Property 2

Captures gapped subsequences

1. My phone calls drop frequently with the iPhone.
2. Great device, but the calls drop too frequently.
Property 2

Captures gapped subsequences

Gap between words = 2
Property 2

Captures gapped subsequences

Gapped subsequences allow:
• redundancy enforcements
• discovery of new sentences
Property 3

Captures collapsible structures

1. Calls drop frequently with the iPhone
2. Calls drop frequently with the Black Berry
Property 3

Captures collapsible structures

Hub-like node

calls drop frequently with the iphone

High redundancy path

Blackberry

High fan-out
Property 3

Captures collapsible structures

"calls drop frequently with the *iphone* and the *black berry*"
Property 3

Captures collapsible structures

- Ideal for collapse & compression
- Can easily be discovered using OG
Step 2a: Generate Candidate Summaries
Generate Candidate Summaries

Repeatedly search the Opinosis-Graph for a Valid Path
Valid Path

- Set of connected nodes

- Has a Valid Start Node (VSN)
  - Natural starting point of a sentence
  - 

- Has a Valid End Node (VEN)
  - Point that completes a sentence
  - Opinosis uses punctuations & conjunctions
Finding Candidate Summaries

, calls drop frequently with the iphone.
Finding Candidate Summaries

, calls drop frequently with the iphone .

VSN? Yes!
Finding Candidate Summaries

, calls drop frequently with the iphone .

VSN

VEN?
Finding Candidate Summaries

, calls drop frequently with the iphone .

VSN

VEN?
Finding Candidate Summaries

, calls drop frequently with the iphone .

VSN

VEN?
Finding Candidate Summaries

, calls drop frequently with the iphone.
Finding Candidate Summaries

, calls drop frequently with the iphone .

VSN  VEN
Finding Candidate Summaries

, calls drop frequently with the iphone.
Finding Candidate Summaries

, calls drop frequently with the iphone .

VSN

Pool of candidate summaries

VEN
Collapsible Structures

- Some paths are collapsible
- We need to identify such paths
Identifying a Collapsible Structure

- Find **collapsible nodes** (hub-like nodes) during traversal

- **Opinosis:** Treat **linking verbs** (e.g. is, are) as collapsible nodes
  - Linking verbs have hub-like properties
  - Commonly used in opinion text
A Collapsible Structure

Collapsible node

the screen is very clear

big
A Collapsible Structure

- Each path is a Valid Path

Diagram:
- Collapsible node
- the screen → is → very → clear
- big
- Valid Path
A Collapsible Structure

- Each path is a **Valid Path**
A Collapsible Structure

Collapsible node

the screen is very clear

big

Anchor
- Common structure
- High redundancy path

Collapsed candidates (CC)
A Collapsible Structure

the screen is very clear

big

COLLAPSE

the screen is very clear and big

Candidate summary
How to collapse?

Different ways to collapse

Linking verbs: concatenate using commas

“The screen is very clear, bright, big”

Better readability:

“The screen is very clear, bright and big”
Use existing Opinosis-Graph:
- Find conjunction that appears most frequently with last collapsed candidate

“The screen is very clear, bright ____ big”
Step 2b: Score Candidate Summaries
Scoring Candidate Summaries

- Various properties can be used
- We introduce 2 types of scoring
Scoring Candidate Summaries

- We want high confidence summaries
  - Select candidates with high redundancy
    - # of sentences sharing same path
    - controlled by gap threshold, $\sigma_{gap}$

+ Summaries with good coverage
  - Select longer sentences
    - level of redundancy $\times$ length of candidate paths
    - Favor longer but redundant sentences
Gap Requirement ($\sigma_{\text{gap}}$)

- Gaps vary between sentences sharing nodes

<table>
<thead>
<tr>
<th>Candidate X</th>
<th>w1</th>
<th>w2</th>
<th>w3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentence 1</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Sentence 2</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Sentence n</td>
<td>m</td>
<td>n</td>
<td></td>
</tr>
</tbody>
</table>

gap         gap
Gap Requirement ($\sigma_{\text{gap}}$)

- $\sigma_{\text{gap}}$ enforces maximum allowed gap between two adjacent nodes

Candidate X

Sentence 1

Sentence 2

$w_1$ $w_2$

1 < $\sigma_{\text{gap}}$

4 $\sigma_{\text{gap}}$

-Lower risk of ill-formed sentences
-Avoids over-estimation of redundancy
Step 3: Final Opinosis Summaries

- Select top 2 scoring candidates that are most dissimilar
Evaluation
Raw Data: User Reviews

- **Hotels**: Tripadvisor.com
- **Products**: Amazon.com
- **Cars**: Edmunds.com
**Dataset Construction**

Ex. iPhone: screen
iPod: battery

51 human generated topics
topics ➔ “opinion seeking queries”

select topic relevant sentences

raw reviews

~100 unordered, topic-related sentences

**Topic 1**
1. sentence 1.....
2. sentence 2.....
3. sentence 3.....
4. sentence 4.....

**Topic 2**
1. sentence 1.....
2. sentence 2.....
3. sentence 3.....
4. sentence 4.....

**Topic 51**
1. sentence 1.....
2. sentence 2.....
3. sentence 3.....
4. sentence 4.....

review document 1
review document 2
review document 51
Input: A review document

Output: A concise abstractive opinion summary

Opinosis Summary:
The screen is big and clear.
Gold Standard

- Human written summaries
  - Concise (~25 words)
  - Focus on summarizing major opinions

- Obtained using Amazon’s Online Workforce (mturk.com)

- ~4 human summaries per topic
Hard to find ‘general’ abstractive summarizer

Use MEAD - Extractive based method
[Radev et al. 2000]

- **Ideal**: selects sentences that capture most essential information
- Select **2 sentences** as the summary
**Evaluation Measures**

- **ROUGE** (rouge-1, rouge-2, rouge-su4)
  - Standard measure for summarization tasks
  - Measures: agreement between system & human summaries

- **Readability Test**
  - Measures: How different *Opinion sums* are compared to human composed summaries?
Human-Human Agreement

- **Estimate**: How much one summary writer agrees with the rest
Human-Human Agreement

<table>
<thead>
<tr>
<th>ROUGE-1</th>
<th>Precision</th>
<th>0.34</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Recall</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>F-score</td>
<td>0.31</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ROUGE-SUM</th>
<th>Precision</th>
<th>0.16</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Recall</td>
<td>0.12</td>
</tr>
</tbody>
</table>

**Manual inspection:** Human summaries semantically similar. Difference in word usage.
Human vs. Opinosis vs. MEAD
Human vs. Opinosis vs. MEAD

Rouge Recall Scores

- **ROUGE-1**
  - HUMAN (17 words): 0.3184
  - OPINOSISbest (15 words): 0.2831
  - MEAD (75 words): 0.4932

- **ROUGE-SU4**
  - HUMAN (17 words): 0.1293
  - OPINOSISbest (15 words): 0.0851
  - MEAD (75 words): 0.2316

Highest recall

Much longer sentences
Overall: Baseline does not do well in generating concise summaries.
Human vs. Opinosis vs. MEAD

Rouge Recall Scores

- **ROUGE-1**
  - HUMAN (17 words): 0.3184
  - OPINOSISbest (15 words): 0.2831
  - MEAD (75 words): 0.4932

- **ROUGE-SU4**
  - HUMAN (17 words): 0.1293
  - OPINOSISbest (15 words): 0.0851
  - MEAD (75 words): 0.2316
Human vs. Opinosis vs. MEAD

Rouge Precision Scores

- HUMAN (17 words): ROUGE-1 = 0.3434, ROUGE-SU4 = 0.3088
- OPINOSISbest (15 words): ROUGE-1 = 0.4482, ROUGE-SU4 = 0.3271
- MEAD (75 words): ROUGE-1 = 0.0916, ROUGE-SU4 = 0.1515
Human vs. Opinosis vs. MEAD

**Rouge Precision Scores**

- ROUGE-1: 0.3434
- ROUGE-SU4: 0.4482
- Performance of Opinosis similar to Human performance
Effect of Gap Threshold ($\sigma_{\text{gap}}$)
Effect of Gap Threshold ($\sigma_{gap}$)

![Graph showing the effect of gap threshold on ROUGE-1 f-score with three different weightings: basic, wt_loglen, and wt_len. The graph plots ROUGE-1 f-score against $\sigma_{gap}$ values from 1 to 5, with $\sigma_{gap}$ increasing from 0.250 to 0.330.]
Effect of Gap Threshold ($\sigma_{gap}$)

- Lowest performance
  - Strict adjacency – disallows redundancies to be captured
Effect of Gap Threshold ($\sigma_{\text{gap}}$)

Jump in performance
More redundancies are captured
Effect of Gap Threshold ($\sigma_{gap}$)

Small improvements afterwards

ROUGE-1 (f-score)

$\sigma_{gap}$

1 2 3 4 5

basic
wt_loglen
wt_len
Effect of Gap Threshold ($\sigma_{\text{gap}}$)

Small improvements afterwards

Graph showing the effect of the gap threshold ($\sigma_{\text{gap}}$) on the ROUGE-1 (f-score) for different conditions:
- Basic
- wt_loglen
- wt_len

The graph illustrates a gradual increase in ROUGE-1 (f-score) with increasing $\sigma_{\text{gap}}$, with notable small improvements afterwards.
Effect of Gap Threshold \((\sigma_{\text{gap}})\)

Small improvements afterwards

Gap too large: **ill formed sentences**

Set \(\sigma_{\text{gap}}\): between 2-5
Compare: Scoring Functions
Compare: Scoring Functions

- **redundancy & path length**
- **only redundancy**

![Graph showing ROUGE-1 (f-score) vs. σ_{gap}](image)

- Orange line: basic
- Green line: wt_loglen
- Red line: wt_len
Compare: Scoring Functions

redundancy & path length performs better

only redundancy

ROUGE-1 (f-score) vs $\sigma_{gap}$

- basic
- wt_loglen
- wt_len
Readability Test
How Readability Test Works

Opinosis Summary
1. sentence 1.....
2. sentence 2.....

Human Summary 1
1. sentence 1.....
2. sentence 2.....
3. sentence X..

Human Summary M
1. sentence 1.....
2. sentence 2.....
3. sentence X.....

Mixed Sentences
1. sentence 1.....
2. sentence 3.....
3. sentence 2.....
4. sentence 4.....
5. sentence 8.....
6. sentence 6.....
7. sentence 7.....
8. sentence 5.....

Pick at most 2 sentences that are least readable
Readability Test

- Assessor often picks Opinosis sentences:
  - Opinosis summaries have issues with readability

- Assessor often picks wrong sentences or makes no picks:
  - Opinosis summaries similar to human summaries
Our Readability Test

- Use summaries from Opinosis\textsubscript{best}

- 565 sentences in total
  - 102 sentences from Opinosis summaries
  - 463 sentences from human summaries

- > 60\% of Opinosis generated sentences are no different from human composed sentences
Summary

- A framework for summarizing highly redundant opinions

- Use graph representation to generate concise abstractive summaries

- General & lightweight: Can be used on any corpus with high redundancies (Twitter comments, Blog comments, etc)
Thank You...