Opinosis:

A Graph Based Approach to Abstractive Summarization of Highly Redundant Opinions

Kavita Ganesan, ChengXiang Zhai, Jiawei Han

University of Illinois @ Urbana Champaign

ILLINOIS

1867

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Opinion Summarization Today…

Customer Reviews

Average Customer Rating

(1,432 customer reviews)

5 star:	(1,040)
4 star:	(227)
3 star:	(63)
2 star:	(25)
1 star:	(77)

Opinion Summary for **iPod**

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

› See and rate all 1 ttributes.

Most Helpful Customer R

3,677 of 3,770 people fou

WARNNING fo 1, 2009

By Hassan B. Bn Hadhra

REAL NAME

Existing methods: Generate structured ratings for an entity

[Lu et al., 2009; Lerman et al., 2009;..]

Amazon Verified Purchase (What's this?)

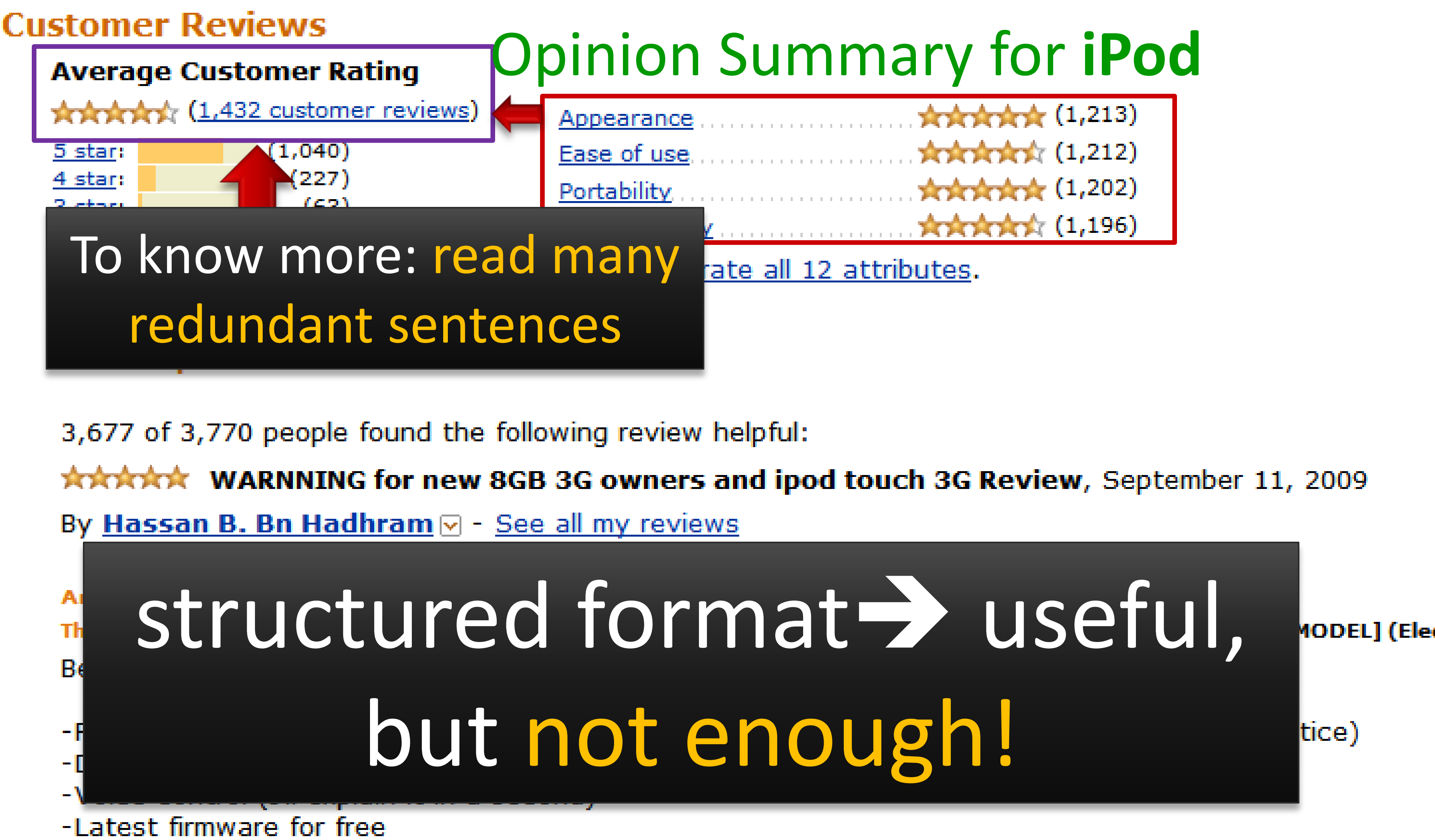
This review is from: Apple iPod touch 8 GB (2nd Generation--with iPhone OS 3.1 Software Installed) [NEWEST MODEL] (Ele

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

Opinion Summarization Today…

Opinion Summary for **iPod**



To know more: read many redundant sentences

structured format→ useful, but not enough!

Ideally, we need….

Supporting textual summary!

A good <u>textual</u> Opinion Summary should...

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

- **Concise**
  - Easily digestible
  - Viewable on smaller screens

- **Readable**
  - Easily understood

An Ideal Summary

"The iPhone's battery lasts long and is cheap but it's bulky.

- Important information summarized
- Concise
- Readable

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:**
  - Generating concise summaries
  - Summarizing highly redundant text
- **Problems**
  - **Bias:** with limit on summary size
    - selected sentence may have missed critical info
  - **Verbose:** May contain irrelevant information
    - not suitable for smaller devices

Extractive Summarization

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

- **But, not suitable for:**
  - Generating concise summaries
  - Summarizing highly redundant text

- **Problems**
  - **Bias:**
    - sele
  - **Verb**
    - not

Extractive ✗
Abstractive ✓

Abstractive Summarization - HARD!!

Existing methods:

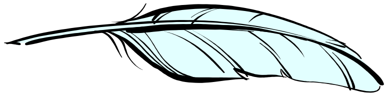
- **Some methods require manual effort**
  [DeJong1982] [Radev and McKeown1998] [Finley and Harabagiu2002]
  - Need to define <u>templates</u> to be filled

- **Some methods rely heavily on 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

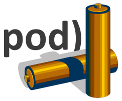
- Uses minimal external knowledge
  - lightweight 

Opinosis: High Level Overview

Opinosis: High Level Overview

Input

Set of sentences:

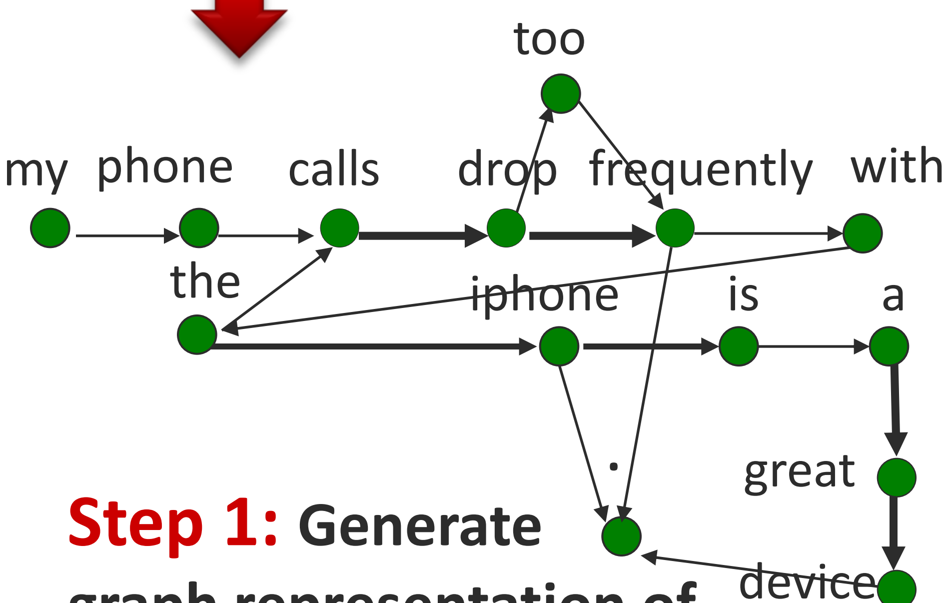
- **Topic** specific (ex. battery life of ipod) 
- **POS** annotated

Opinosis: High Level Overview

Set of sentences:

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

Input

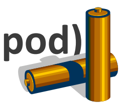


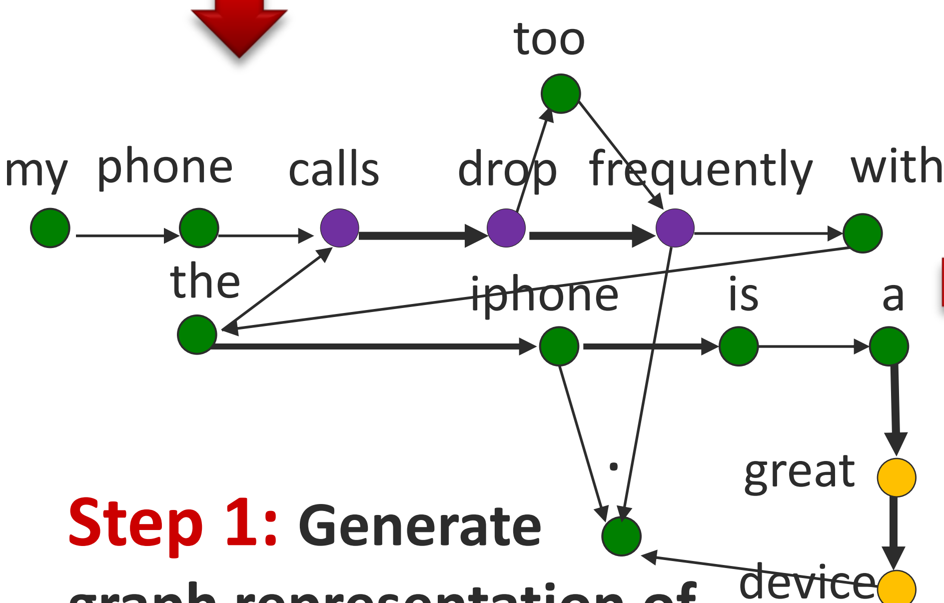
Step 1: **Generate <u>graph representation</u> of sentences (Opinosis-Graph)**

Opinosis: High Level Overview

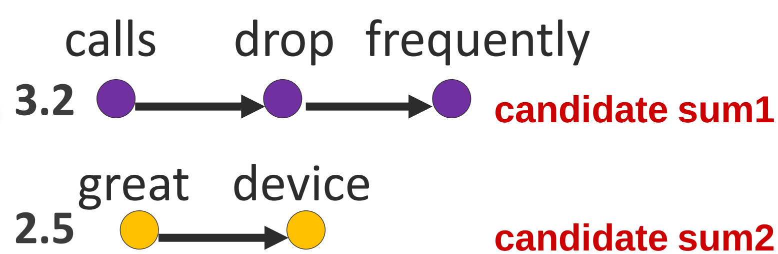
Input

Set of sentences:

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



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



Step 2: Find promising paths (candidate summaries) & score these candidates

Opinosis: High Level Overview

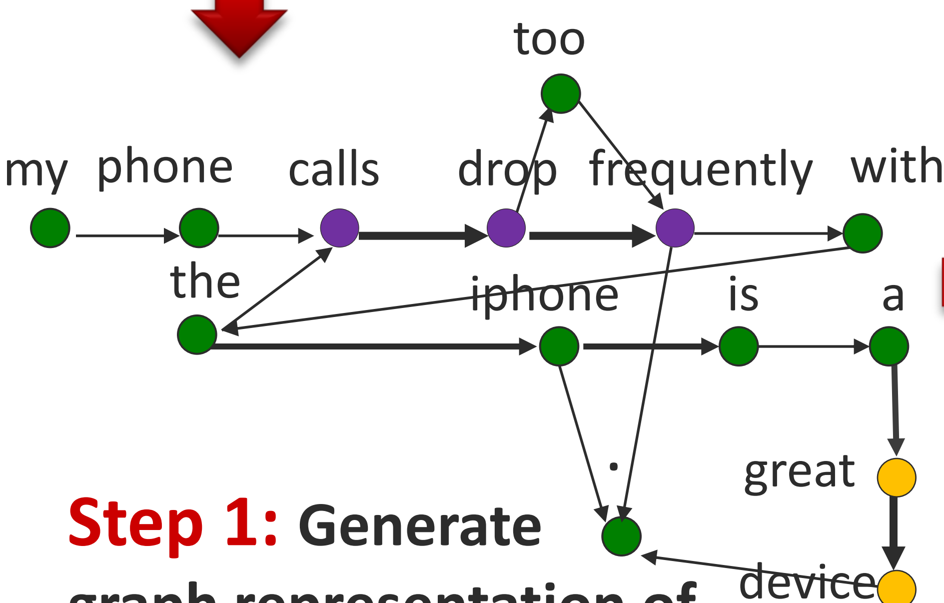
Input

Set of sentences:

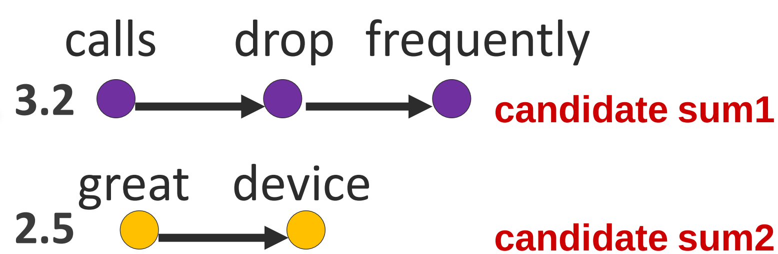
- **Topic specific** (ex. battery li
- **POS annotated**

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

Step 3: Select <u>top scoring candidates</u> as final summary



Step 1: Generate <u>graph representation</u> of sentences (**Opinosis-Graph**)



Step 2: Find <u>promising paths</u> (candidate summaries) & <u>score these candidates</u>

Step 1: Building the Opinosis-Graph

Building 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.*

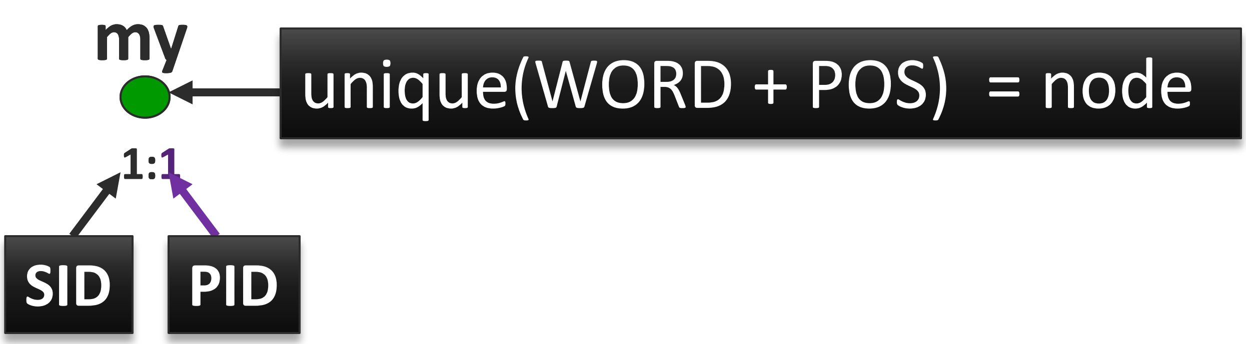
- Opinosis-Graph is empty

Building Opinosis-Graph

1. *My phone calls drop frequently with the iPhone.*

Building Opinosis-Graph

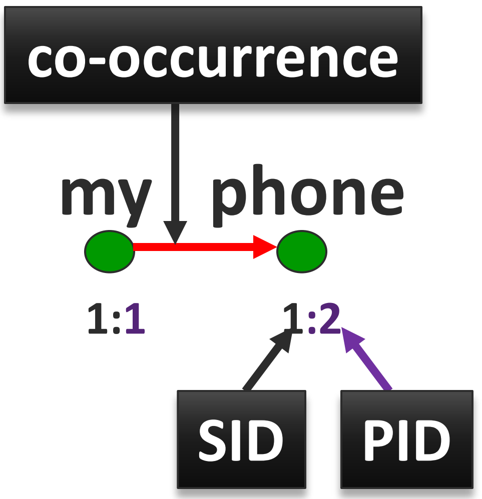
1. *My phone calls drop frequently with the iPhone.*



Positional Reference Information

Building Opinosis-Graph

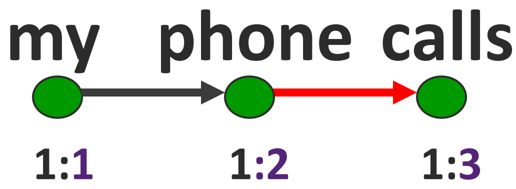
1. *My phone calls drop frequently with the iPhone.*



Positional Reference Information

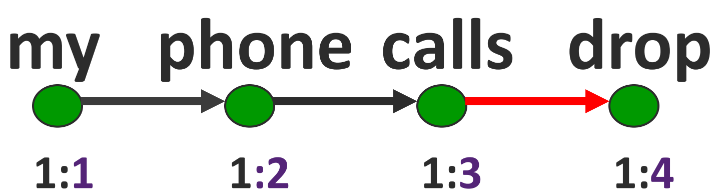
Building Opinosis-Graph

1. *My phone calls drop frequently with the iPhone.*



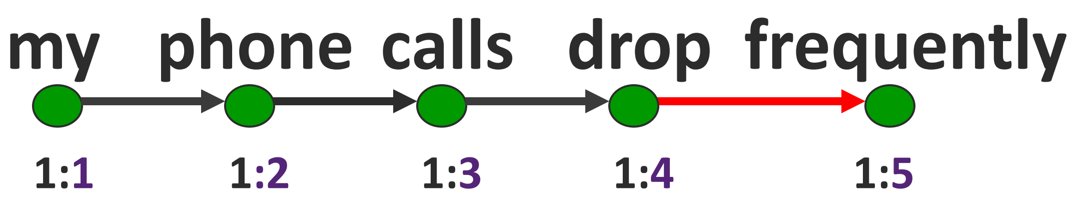
Building Opinosis-Graph

1. *My phone calls drop frequently with the iPhone.*



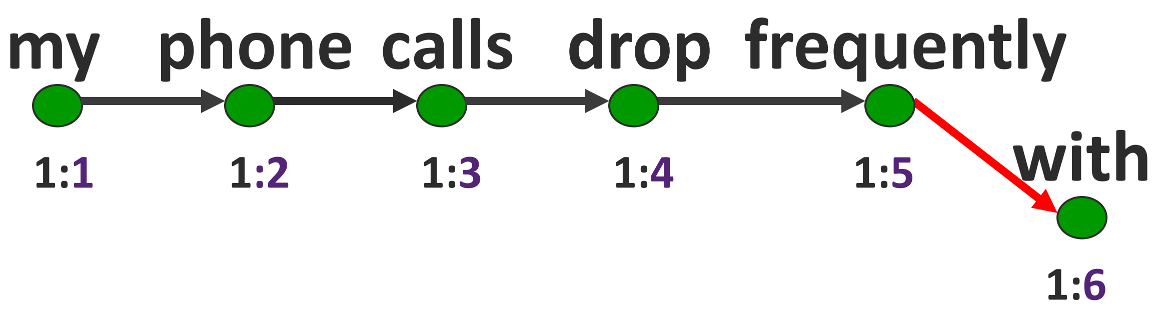
Building Opinosis-Graph

1. *My phone calls drop frequently with the iPhone.*



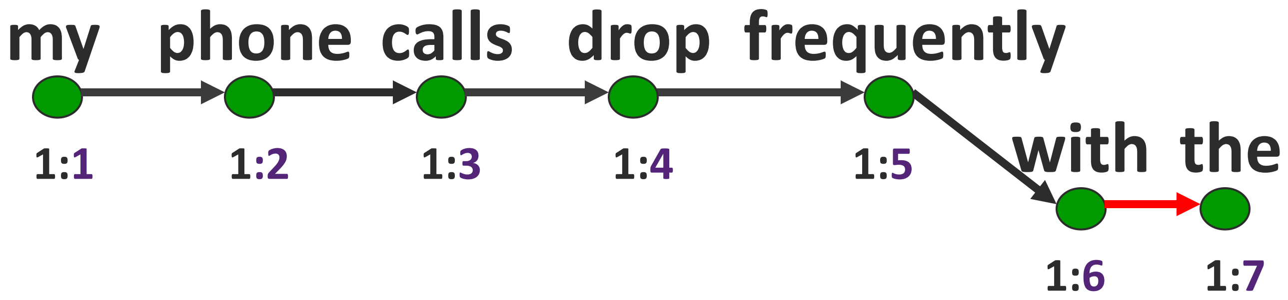
Building Opinosis-Graph

1. *My phone calls drop frequently with the iPhone.*



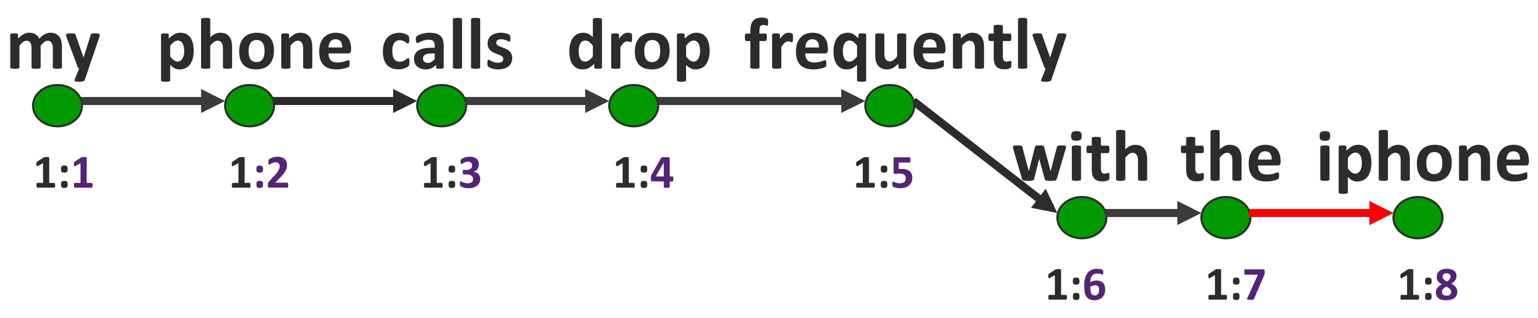
Building Opinosis-Graph

1. *My phone calls drop frequently with the iPhone.*



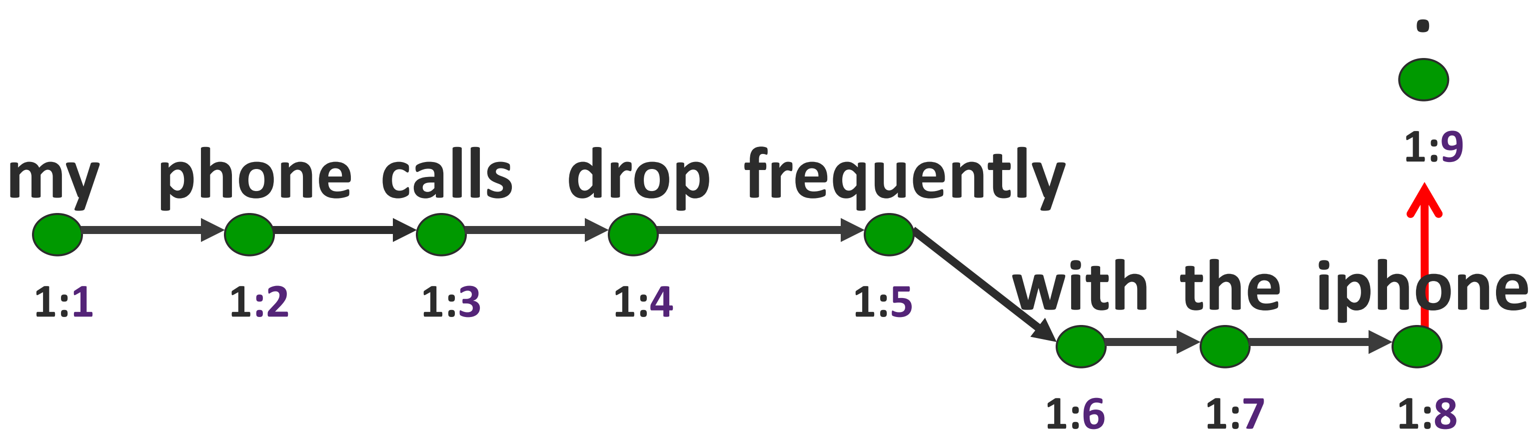
Building Opinosis-Graph

1. *My phone calls drop frequently with the iPhone.*



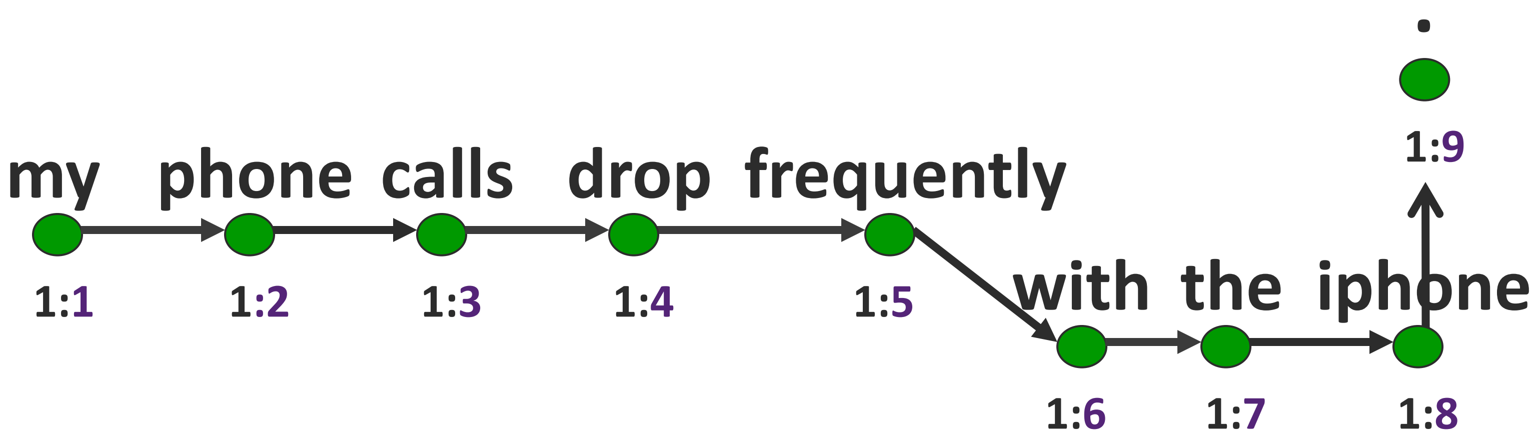
Building Opinosis-Graph

1. *My phone calls drop frequently with the iPhone.*



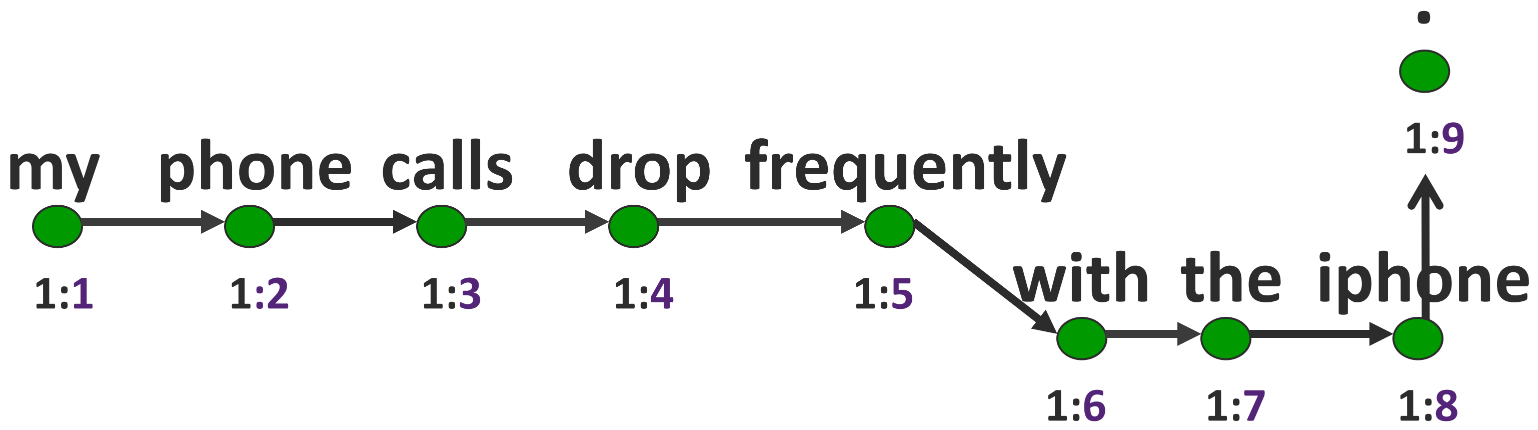
Building Opinosis-Graph

1. *My phone calls drop frequently with the iPhone.*



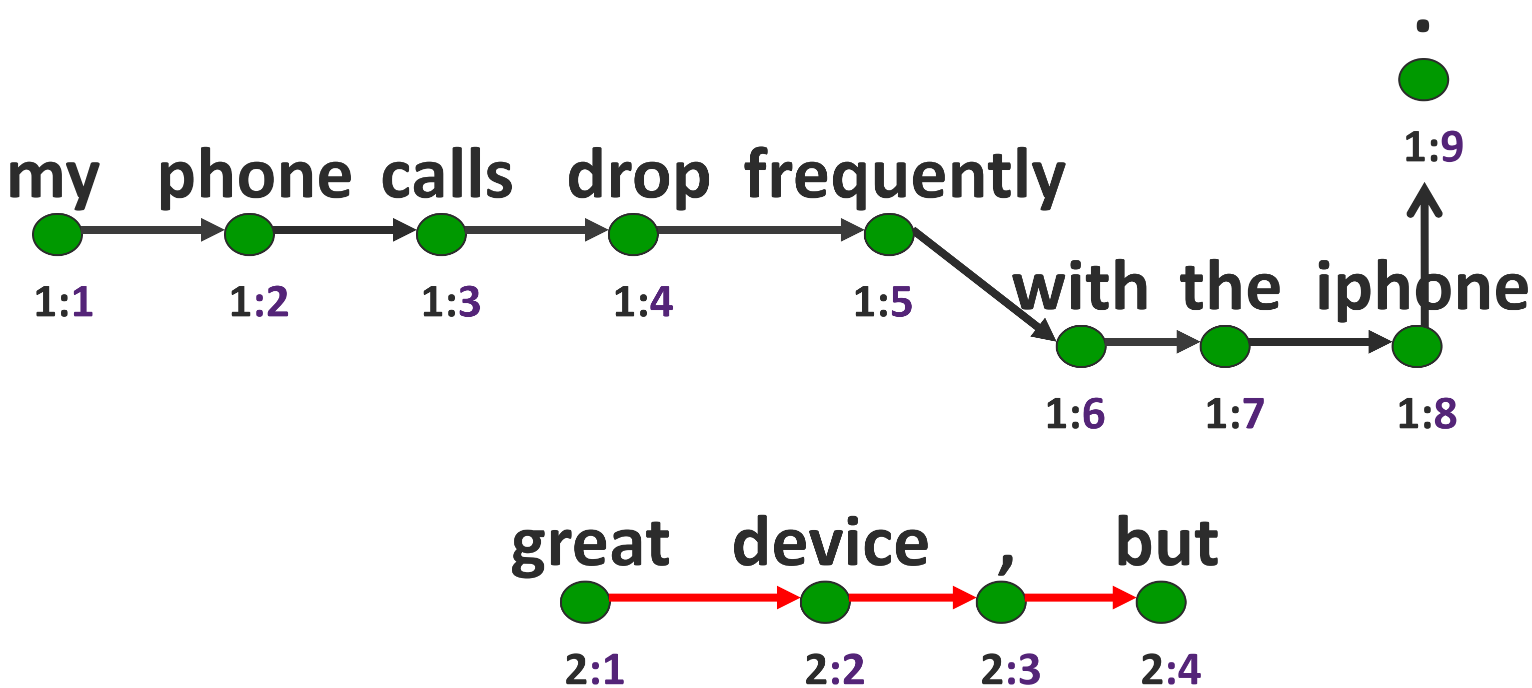
Building Opinosis-Graph

2. *Great device, but the calls drop too frequently.*



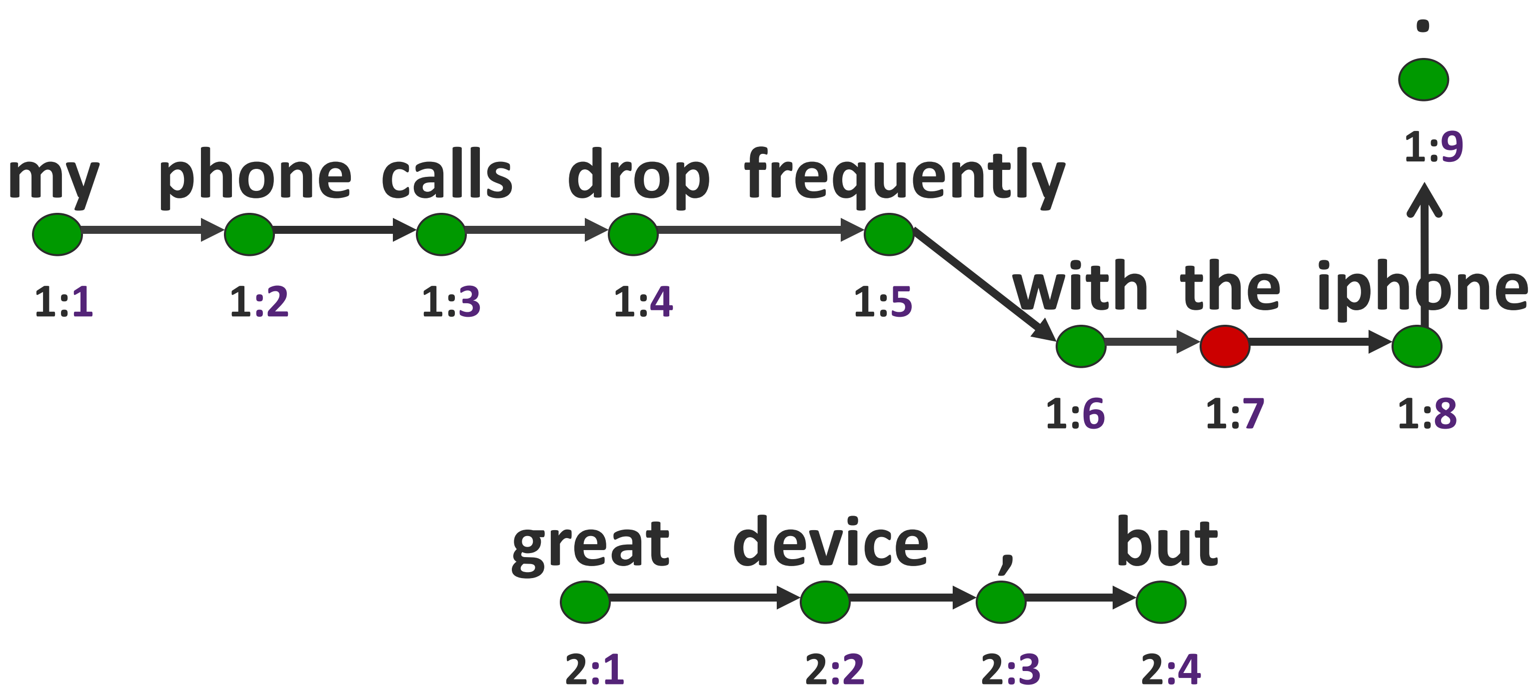
Building Opinosis-Graph

2. *Great device, but the calls drop too frequently.*



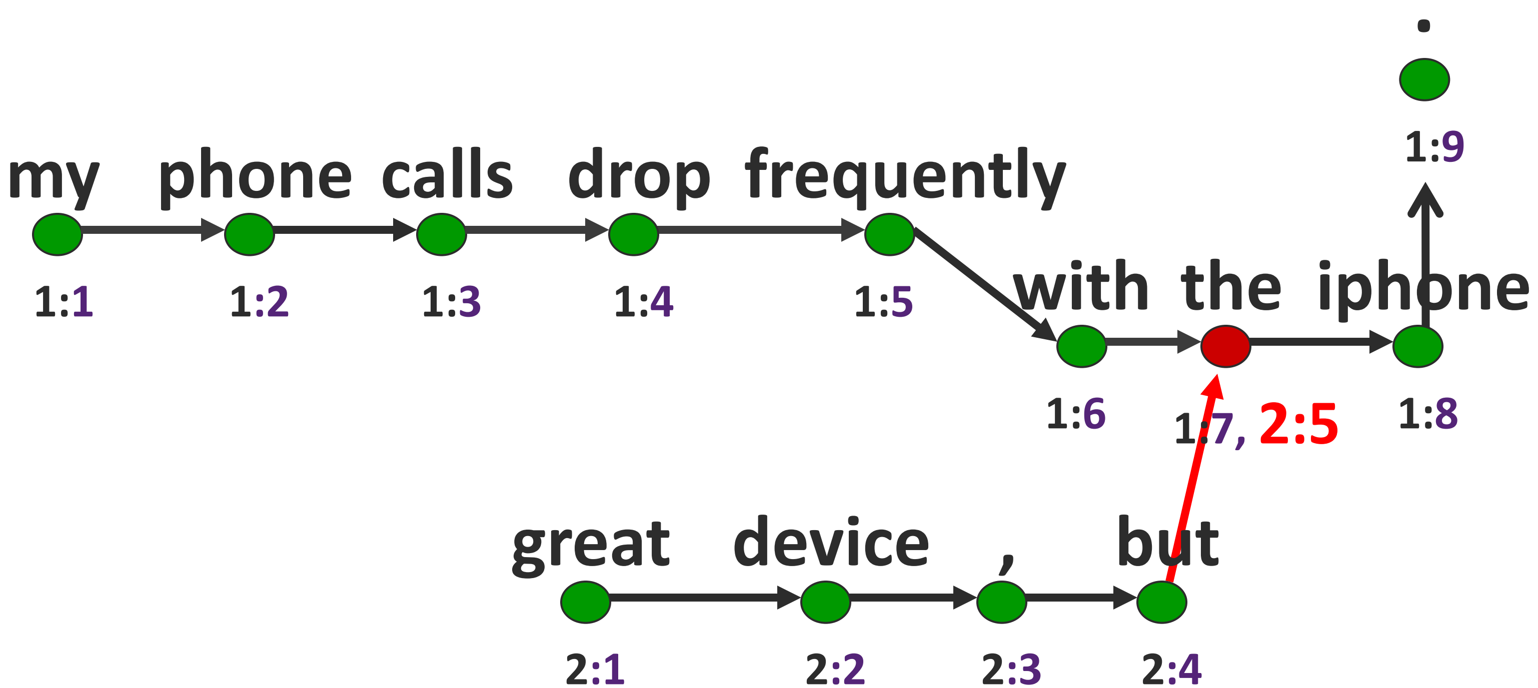
Building Opinosis-Graph

2. *Great device, but the calls drop too frequently.*



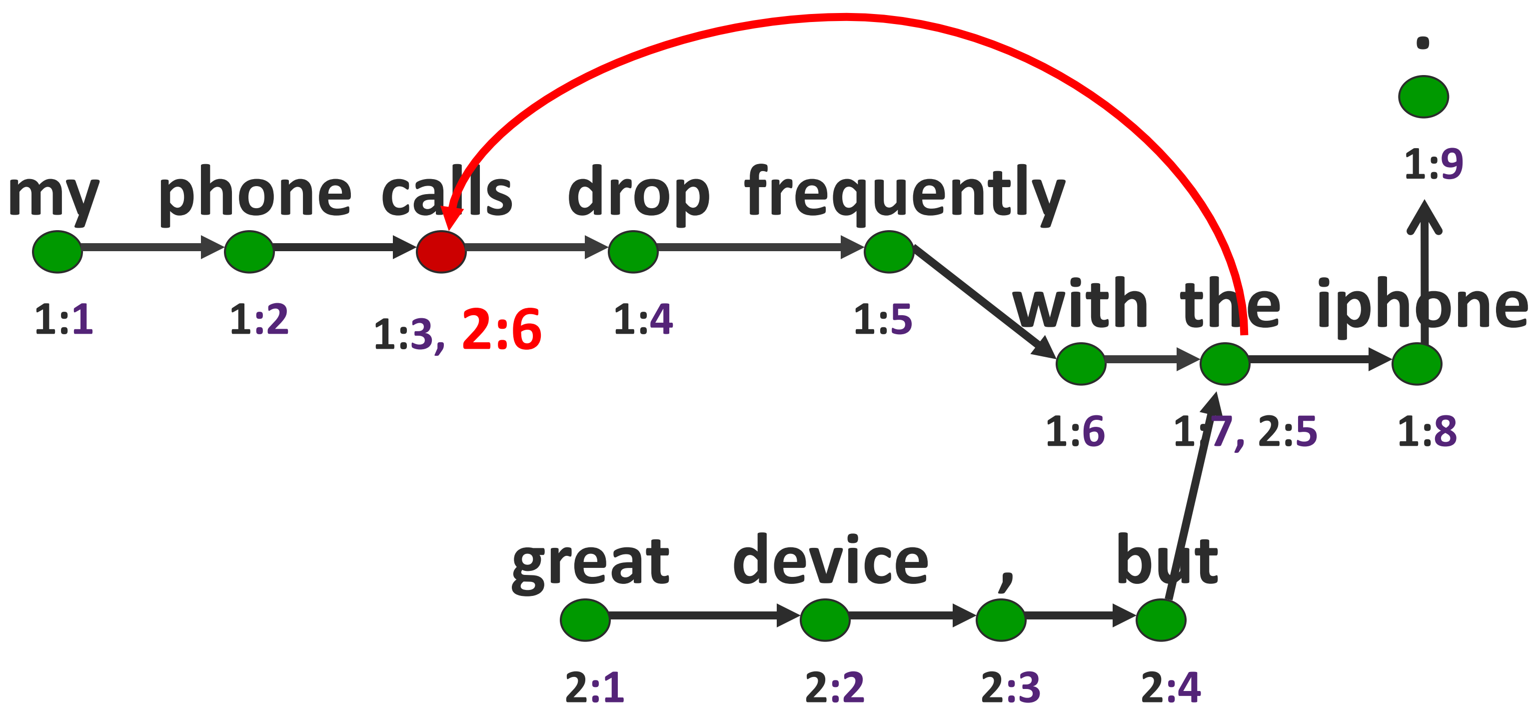
Building Opinosis-Graph

2. *Great device, but the calls drop too frequently.*



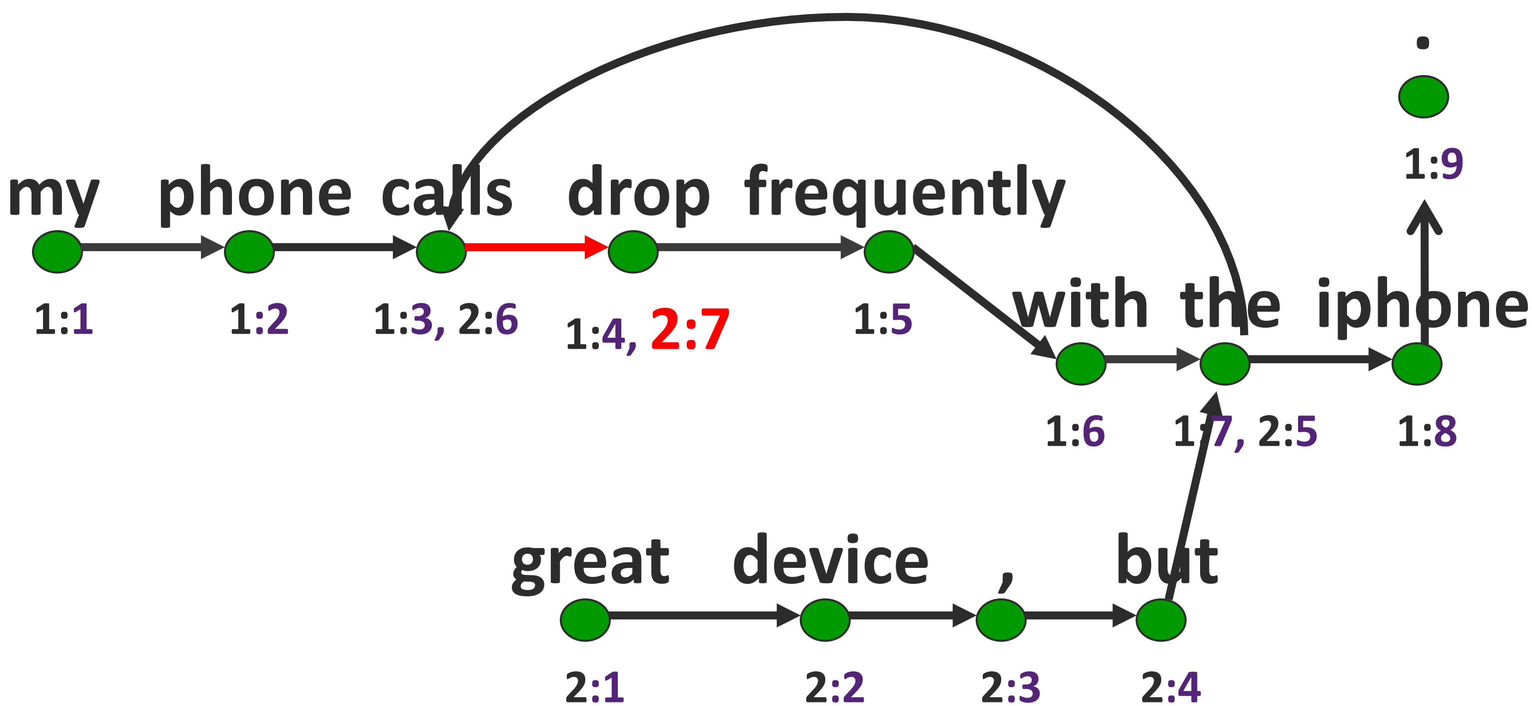
Building Opinosis-Graph

2. *Great device, but the calls drop too frequently.*



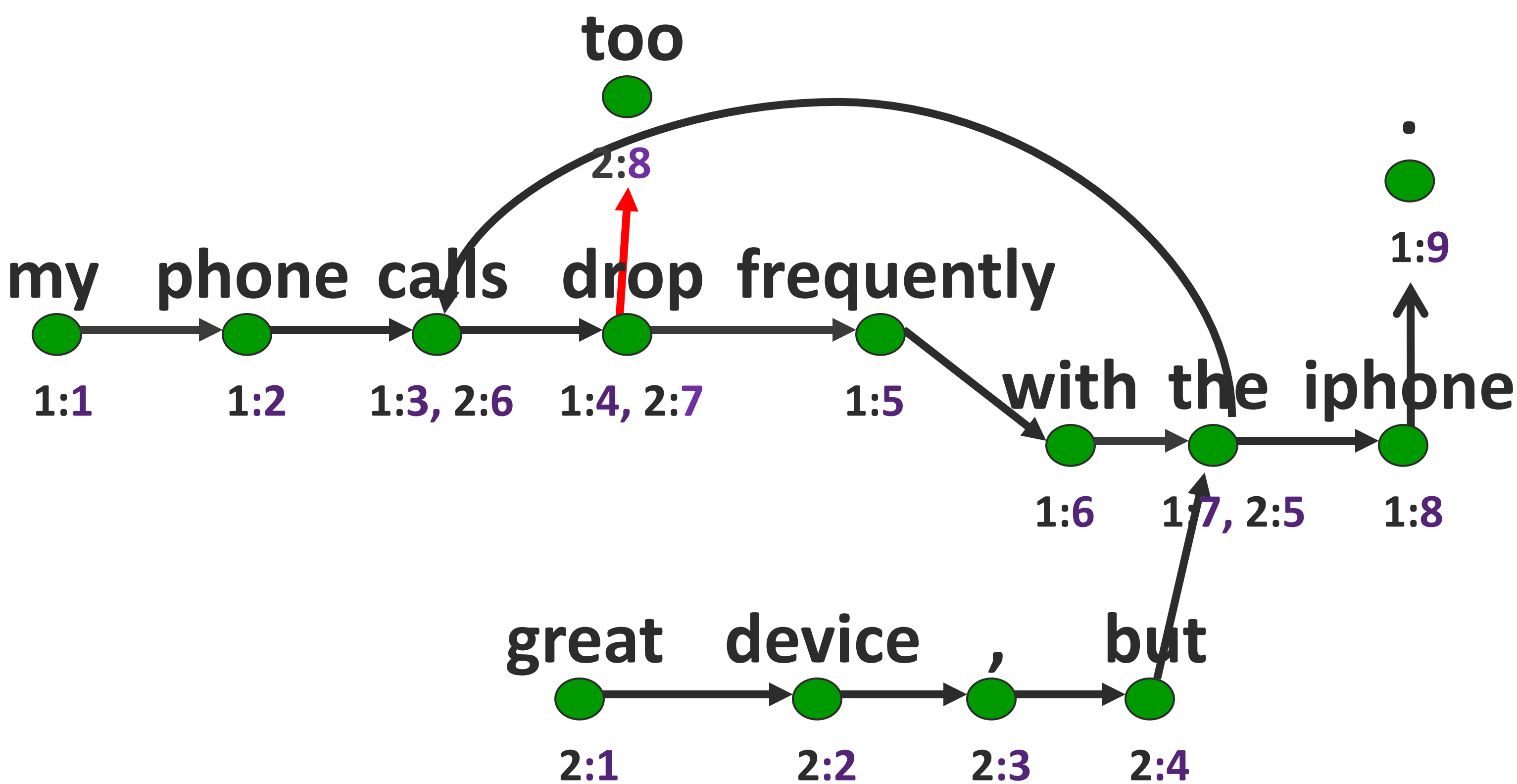
Building Opinosis-Graph

2. Great device, but the calls drop too frequently.



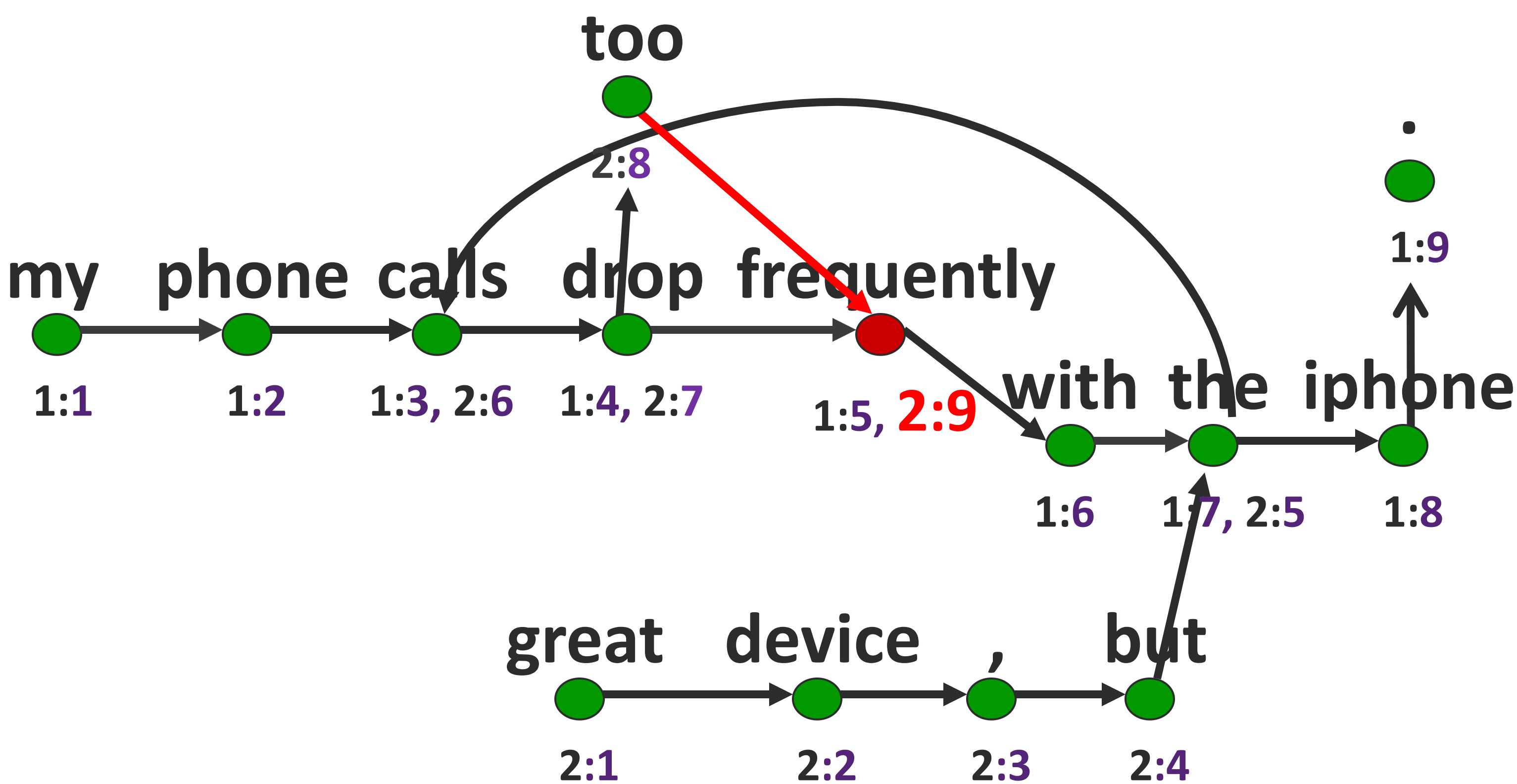
Building Opinosis-Graph

2. Great device, but the calls drop too frequently.



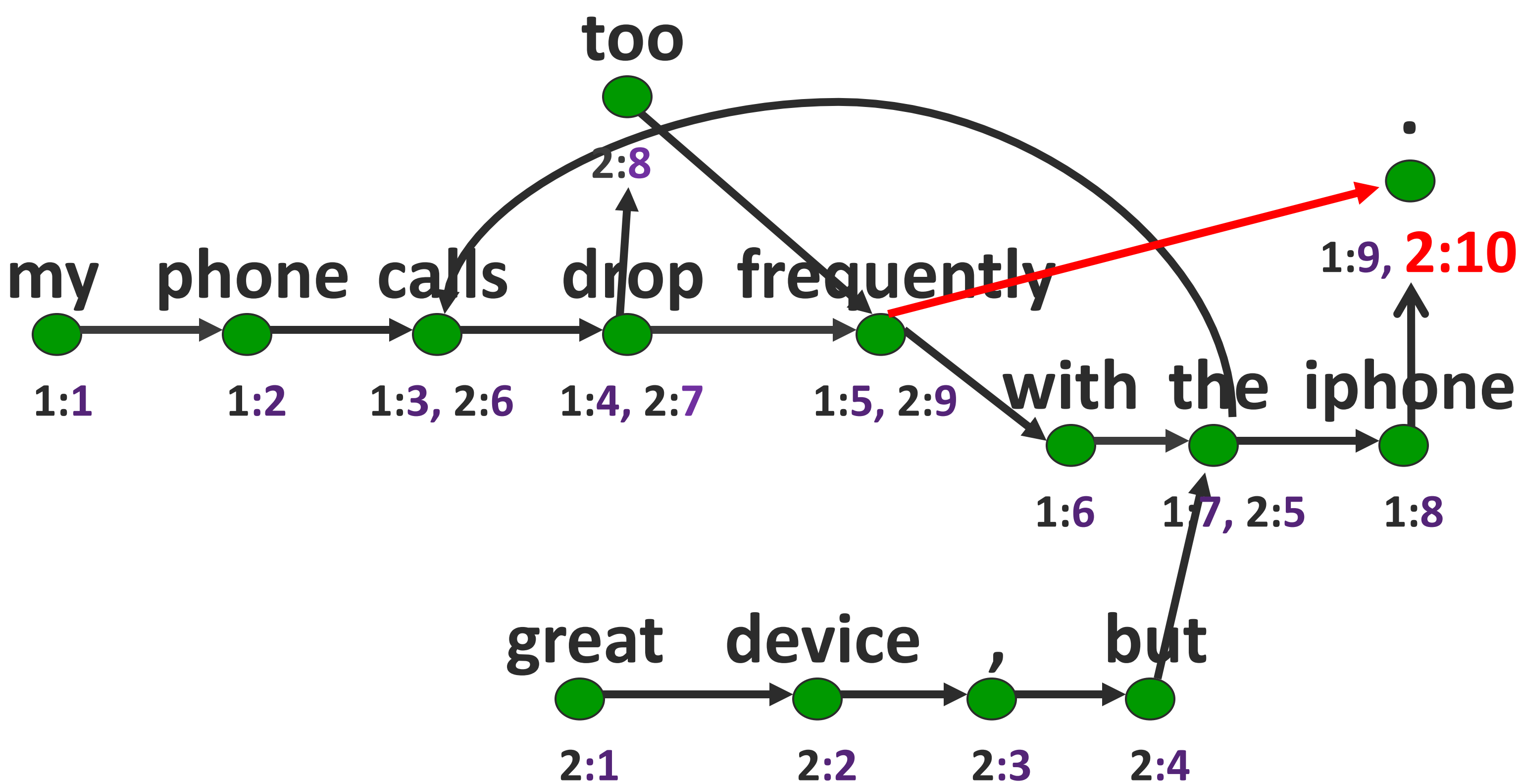
Building Opinosis-Graph

2. *Great device, but the calls drop too frequently.*



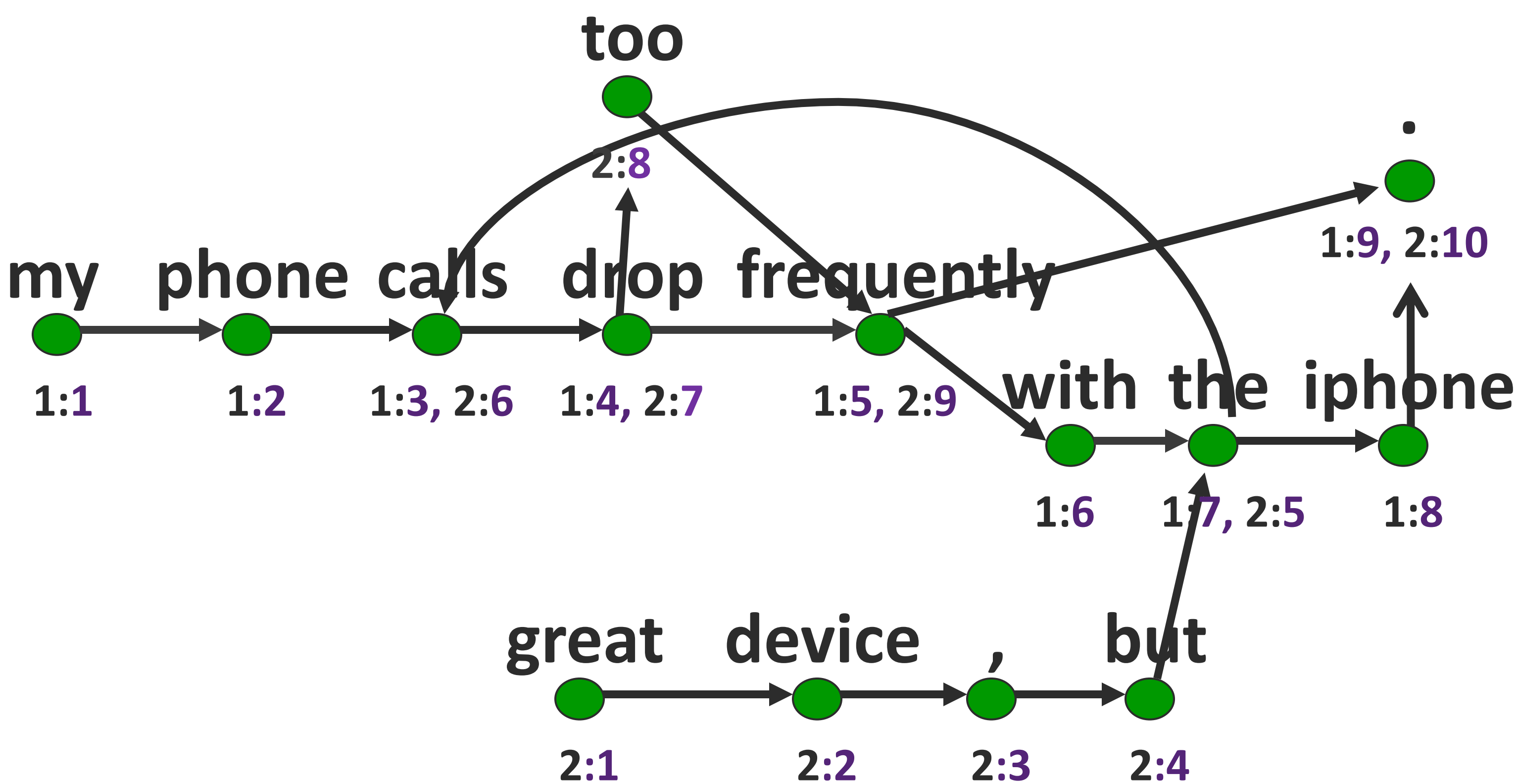
Building Opinosis-Graph

2. Great device, but the calls drop too frequently.



Building Opinosis-Graph

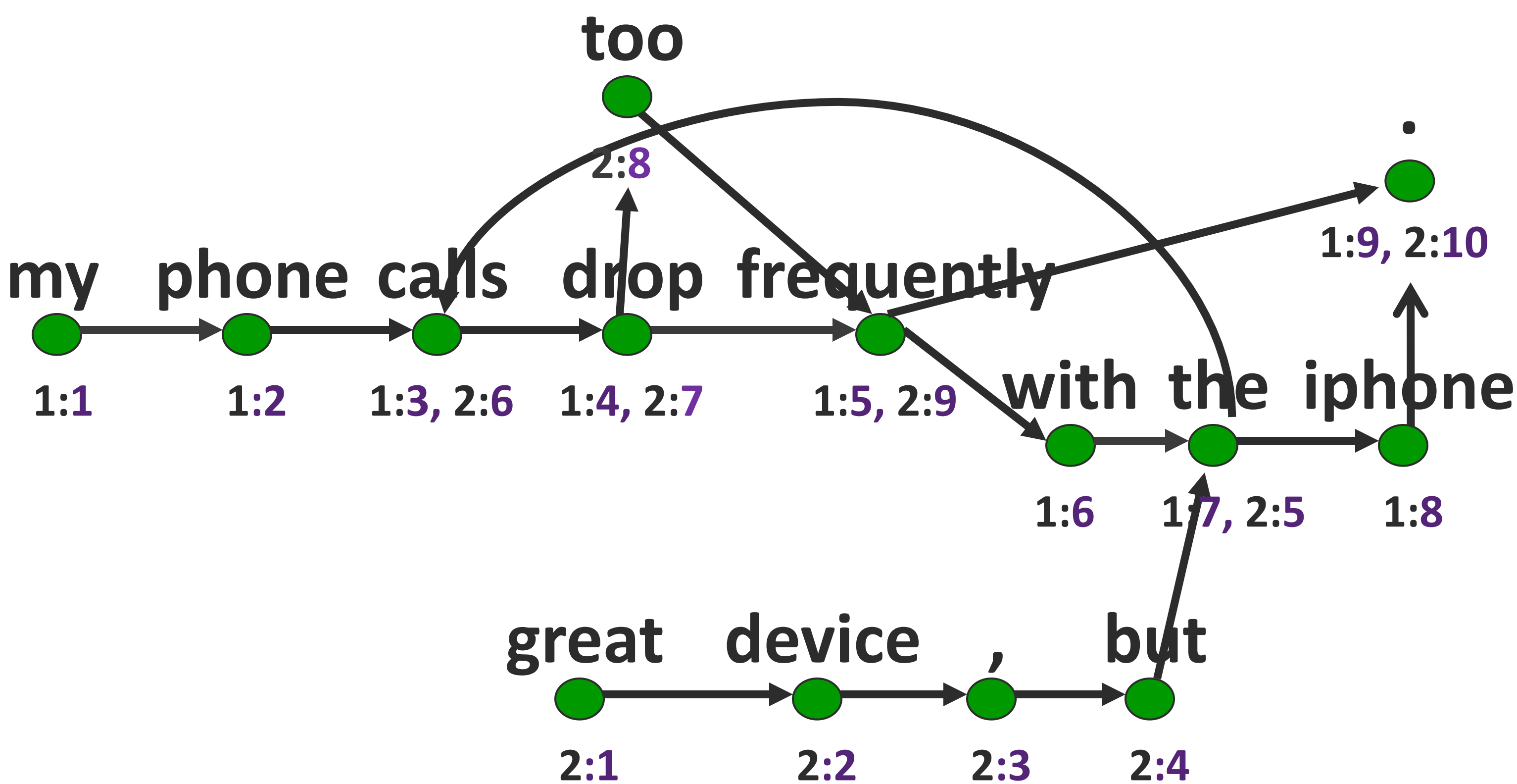
Graph is now ready for Step 2!



3 Important Properties of the Opinosis-Graph

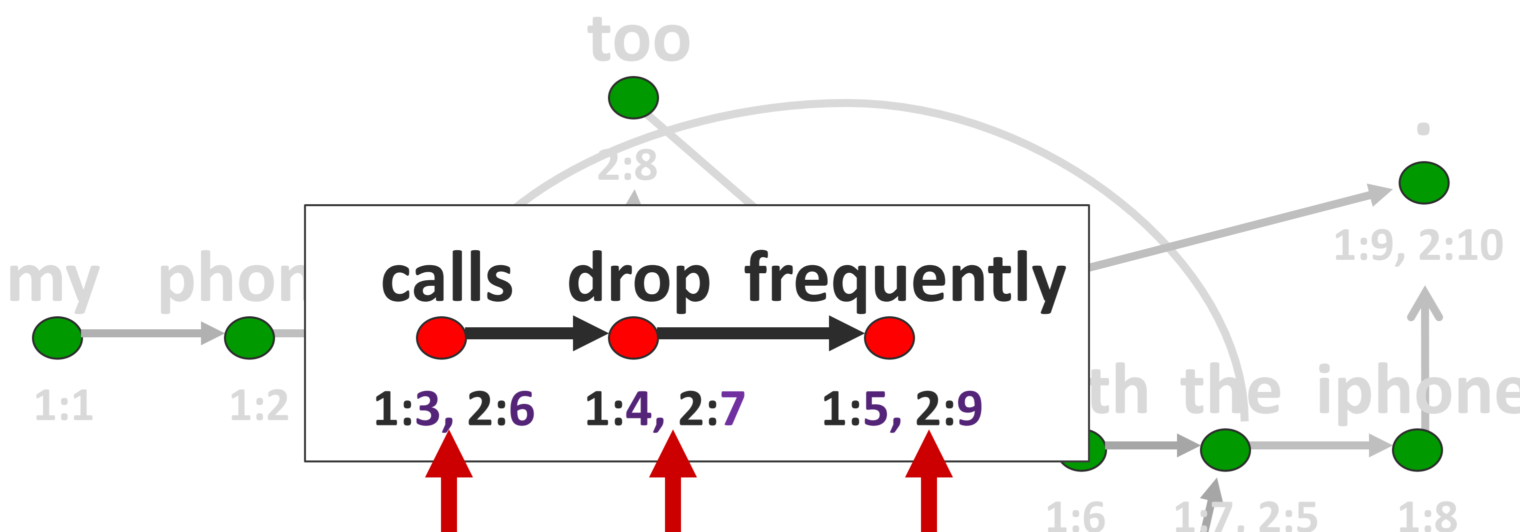
Property 1

Naturally captures redundancies



Property 1

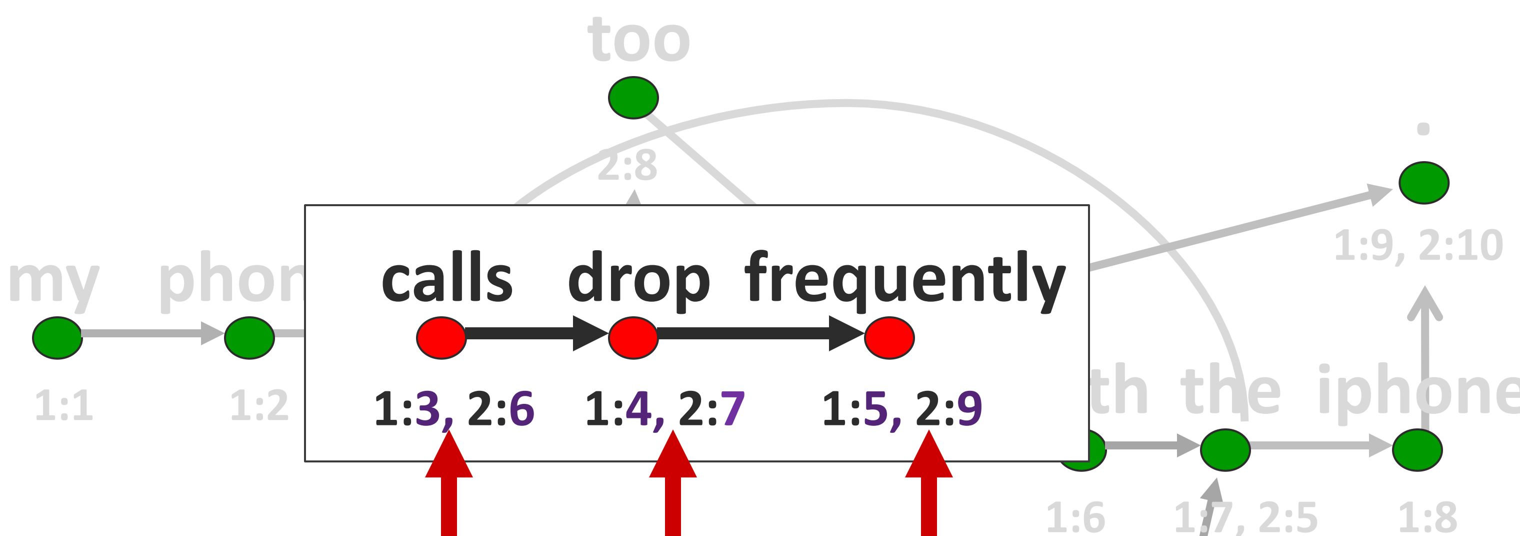
Naturally captures redundancies



Path shared by 2 sentences naturally captured by nodes

Property 1

Naturally captures redundancies



Easily discover redundancies for high confidence summaries

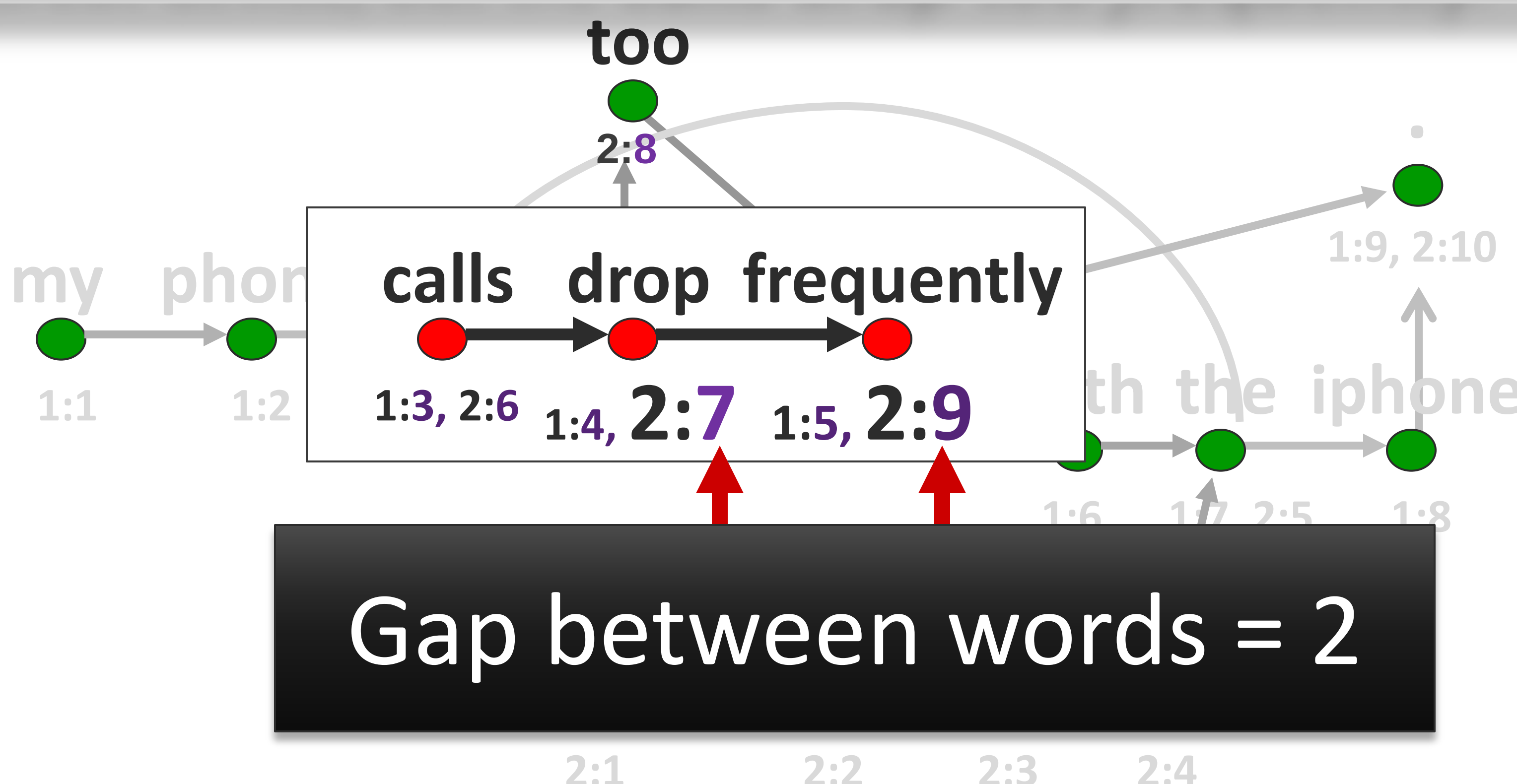
Property 2

Captures gapped subsequences

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

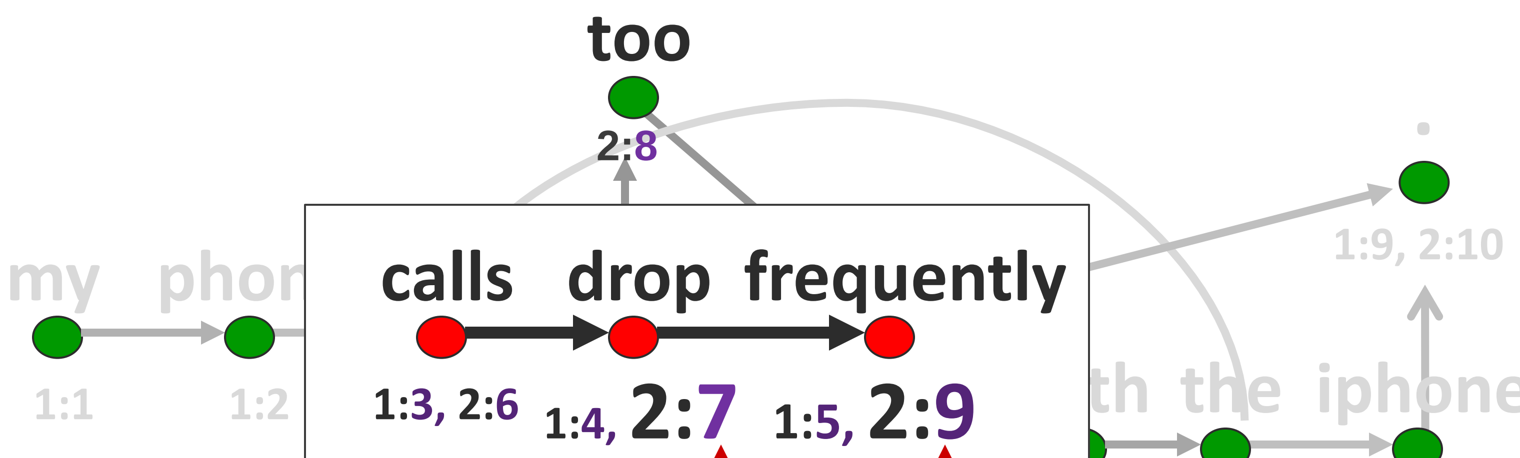
Property 2

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



Property 2

Captures gapped subsequences



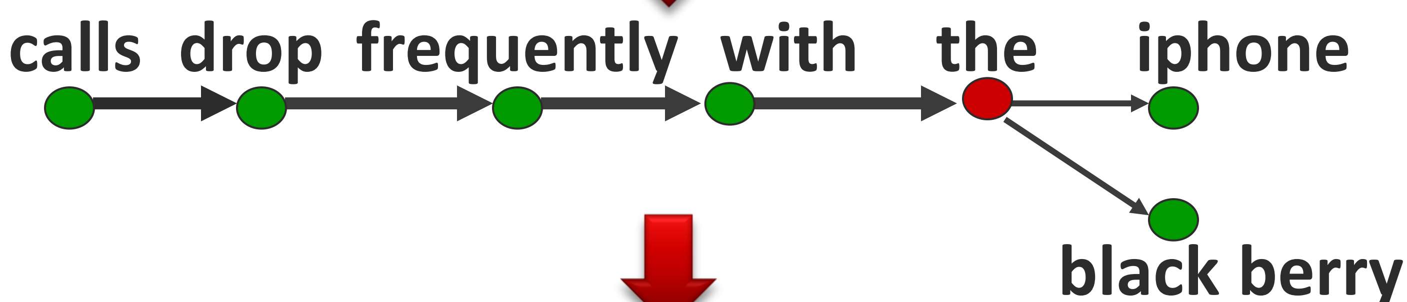
Gapped subsequences allow:

- 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***

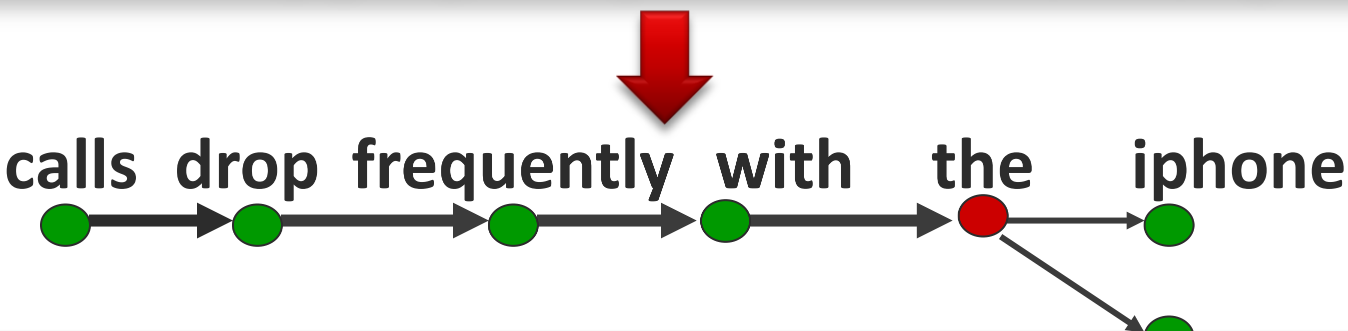


Calls drop frequently with the iPhone and Black Berry

Property 3

Captures collapsible structures

1. *Calls drop frequently with the **iPhone***
2. *Calls drop frequently with the **Black Berry***



- Can easily be discovered using OG
- Ideal for collapse & compression

Step 2a: Generate Candidate Summaries

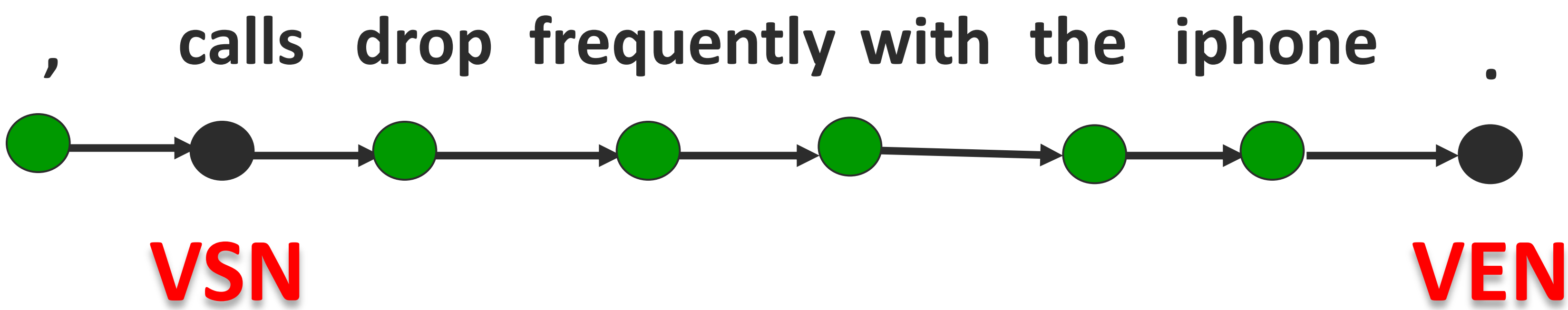
Generate Candidate Summaries

Repeatedly search the Opinosis-Graph for a Valid Path

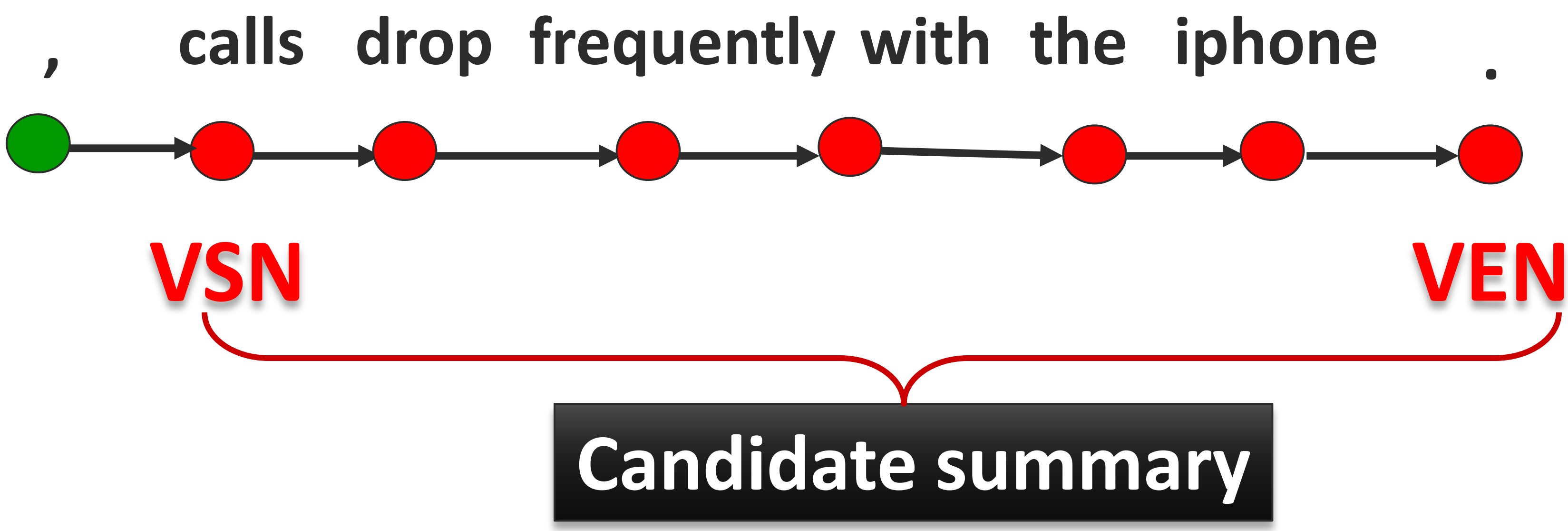
Valid Path

- **Set of connected nodes**
- **Has a Valid Start Node (VSN)**
  - Natural starting point of a sentence
  - Opinosis uses avg. positional information
- **Has a Valid End Node (VEN)**
  - Point that completes a sentence
  - Opinosis uses <u>punctuations</u> & <u>conjunctions</u>

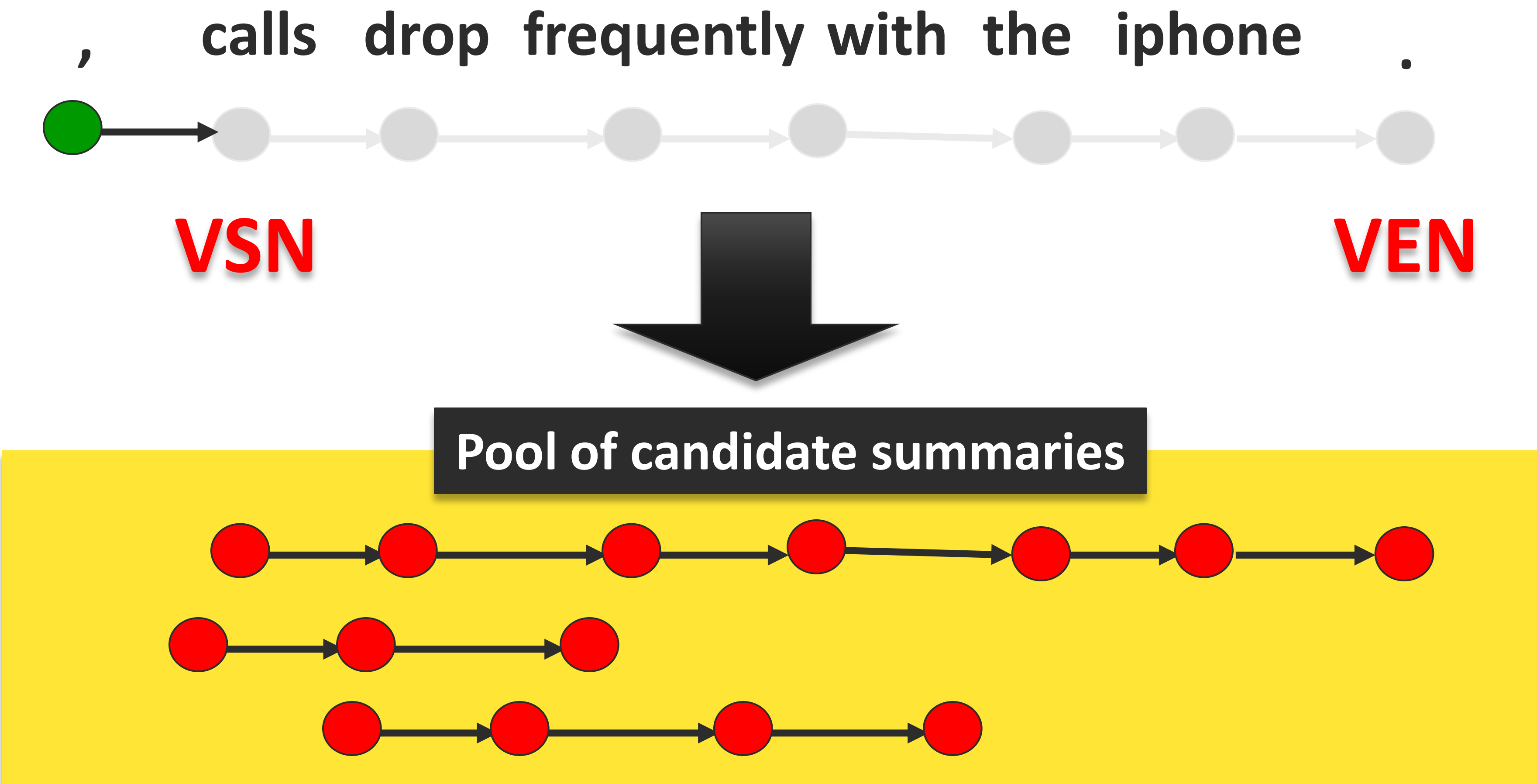
Finding Candidate Summaries



Finding Candidate Summaries



Finding Candidate Summaries



Collapsible Structures

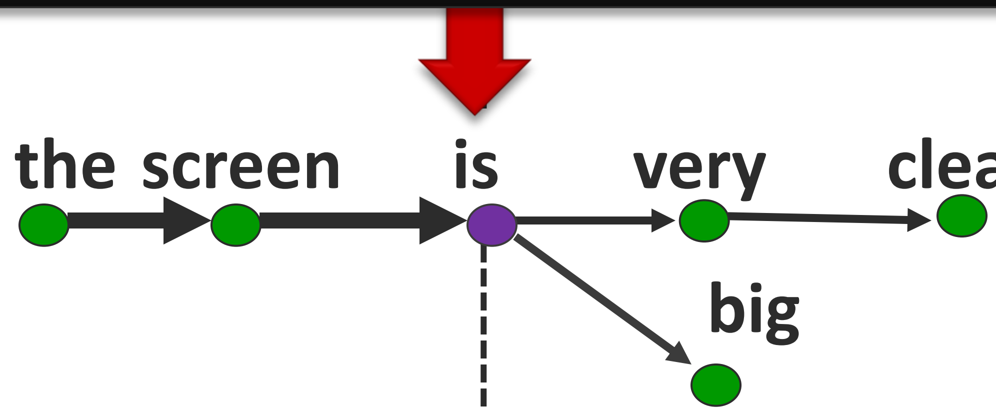
- **Some paths are collapsible**

- **Identify such paths through a collapsible node**

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

A Collapsible Structure

linking verb = collapsible node



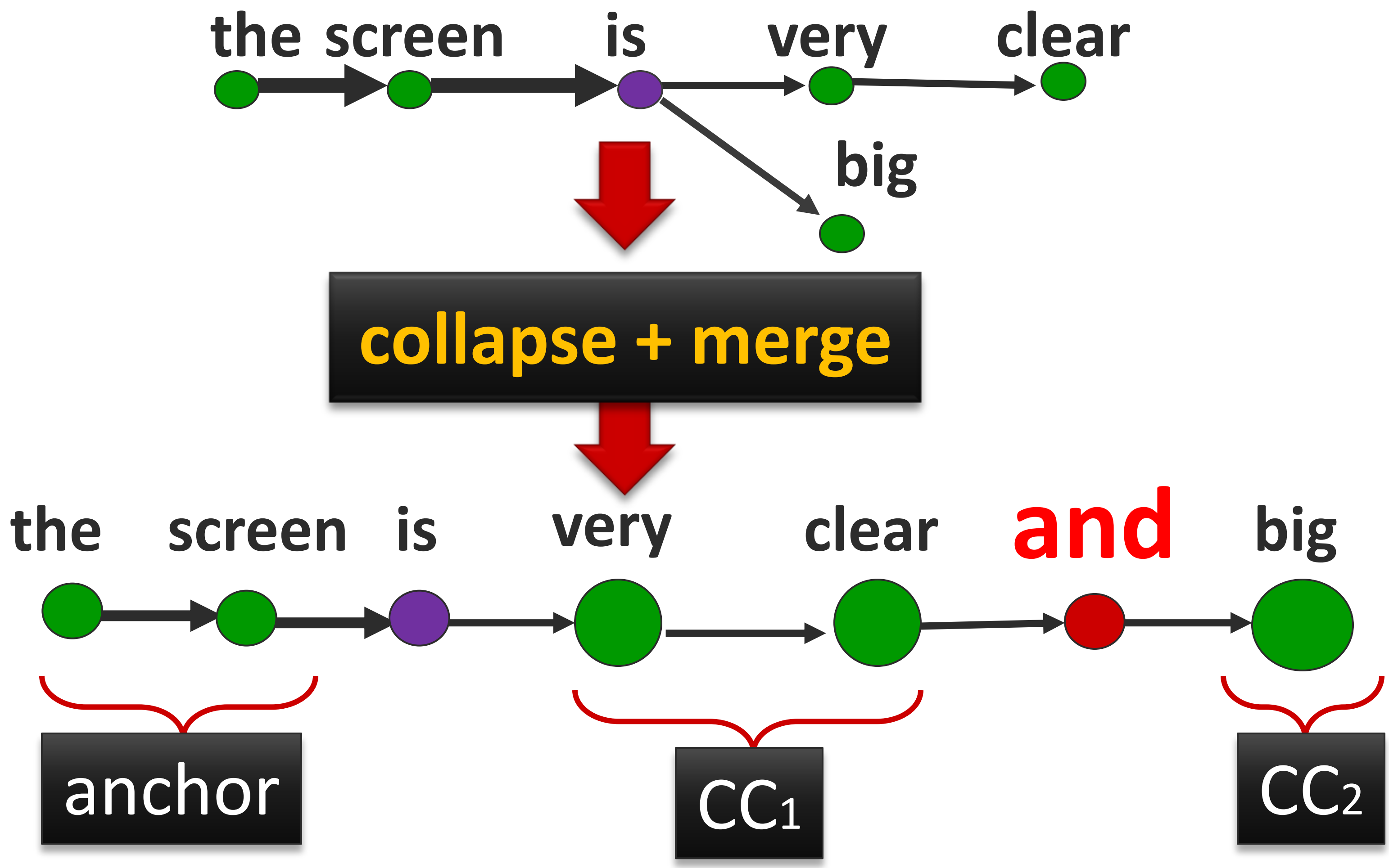
anchor

- *Common structure*
- *High redundancy path*

collapsed candidates (CC)

- *Subgraphs to be merged*

A Collapsible Structure



How to Merge Structures?

- **CC after linking verbs:** concatenate using commas

  "The screen is <u>very clear</u>, <u>bright</u>, <u>big</u>"

  CC_1 CC_2 CC_3

- **Better readability:**

  "The screen is <u>very clear</u>, <u>bright</u> ***and*** *<u>big</u>"*

Find last connector using hints from OG

Step 2b: Score Candidate Summaries

Scoring Options

Type 1: High confidence summaries

- Select candidates with <u>high redundancy</u>
  - # of sentences sharing same path
  - controlled by gap threshold, σ_{gap}

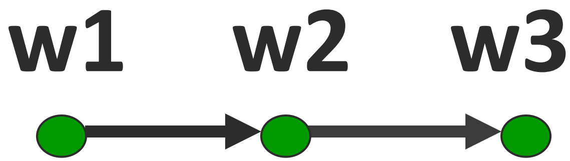
Type 2: + Good coverage

- Select <u>longer candidates</u>
- redundancy * length of candidate paths
  - Favor longer but redundant candidates

Gap Threshold (σ_{gap})

- Gaps vary between sentences sharing nodes

Candidate X



Sentence 1	1	2
Sentence 2	4	4
Sentence X	m	n
	gap	gap

Gap Threshold (σ_{gap})

- σ_{gap} enforces maximum allowed gap between two adjacent nodes

w1 → w2

Candidate X

Sentence 1 1 < σ_{gap} ✓

~~Sentence 2 4 > σ_{gap}~~ ✗

-Lower risk of ill-formed sentences

-Avoids over-estimation of redundancy

Step 3: Final Opinosis Summaries

- **After candidate scoring:**
  - Select top 2 scoring candidates
  - Most dissimilar candidates

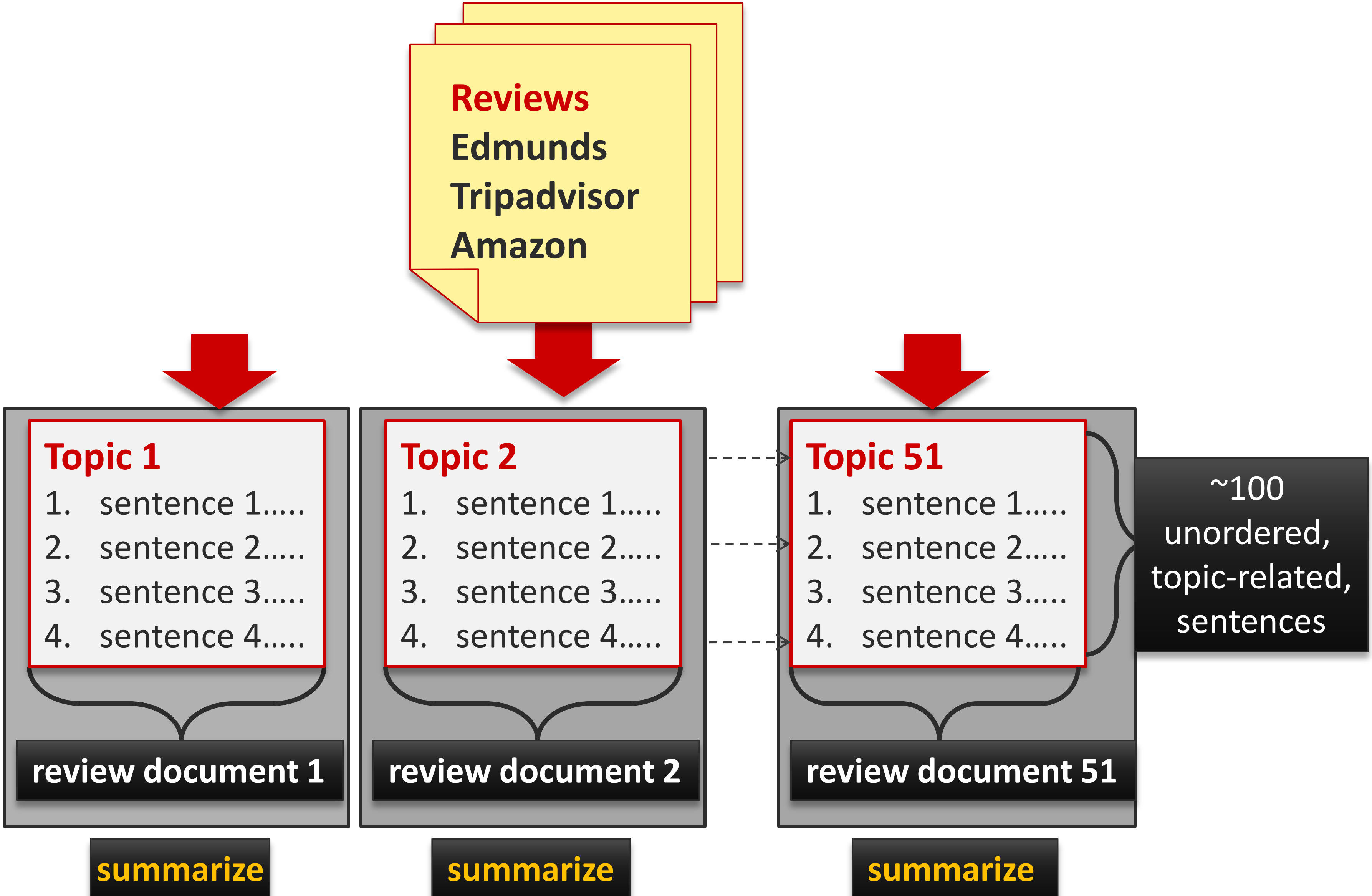
Evaluation

Data

User Reviews:

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

Summarization Task



Gold Standard

- **Human composed summaries**
  - Concise (<25 words)
  - Focus on summarizing major opinions
  - ~4 human summaries per topic

Baseline

- **Hard to find 'general' abstractive summarizer**

- **Use MEAD - Extractive based method** [Radev et al.2000]
  - Select **2 sentences** as the summary

Evaluation Measures

- **ROUGE** (rouge-1, rouge-2, rouge-su4)
  - Standard measure for summarization tasks

- **Readability Test**
  - Measures: How different Opinosis summaries are compared to human composed summaries?

Results

Human Performance

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

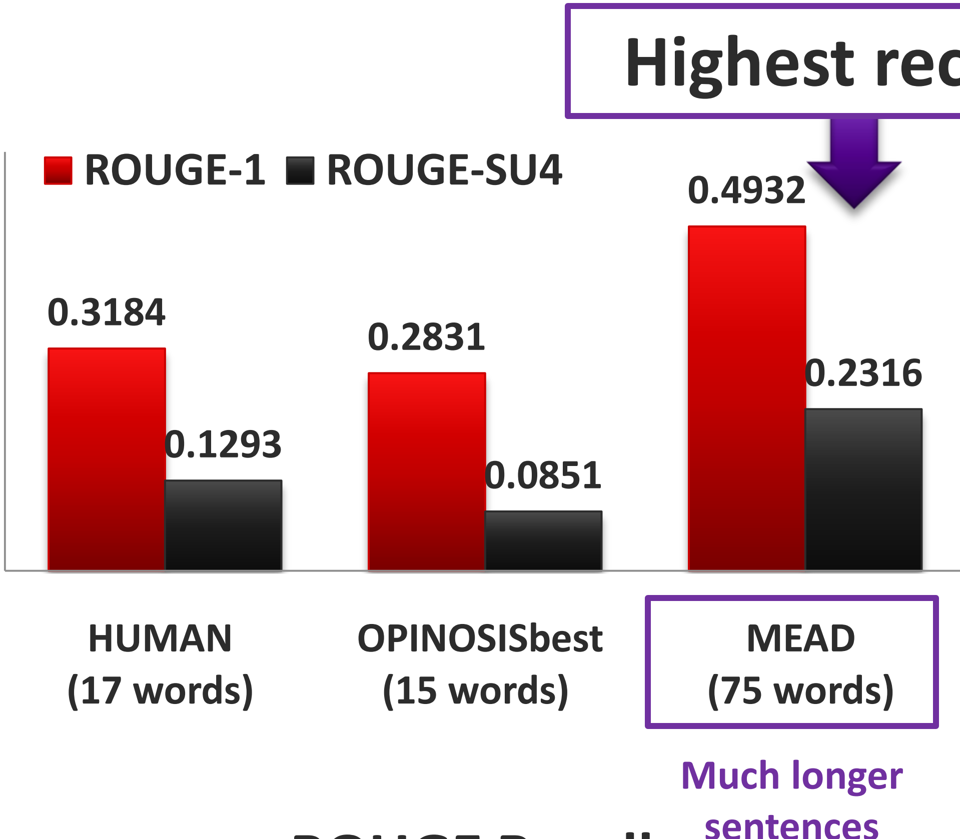
Human Performance

ROUGE Scores		
ROUGE-1	Precision	0.34
	Recall	0.32
	F-score	0.31
ROUGE-SU4	Precision	0.16
	Recall	0.13
	F-score	0.11

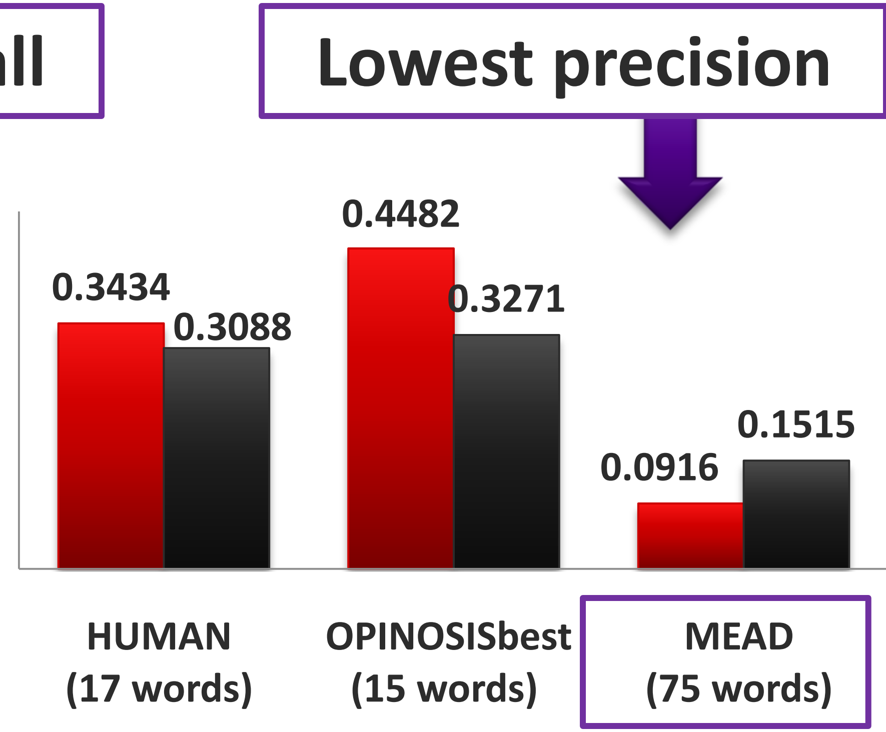
Human summaries - semantically similar. Slight difference in word usage.

Human vs. Opinosis vs. MEAD

Human vs. Opinosis vs. MEAD

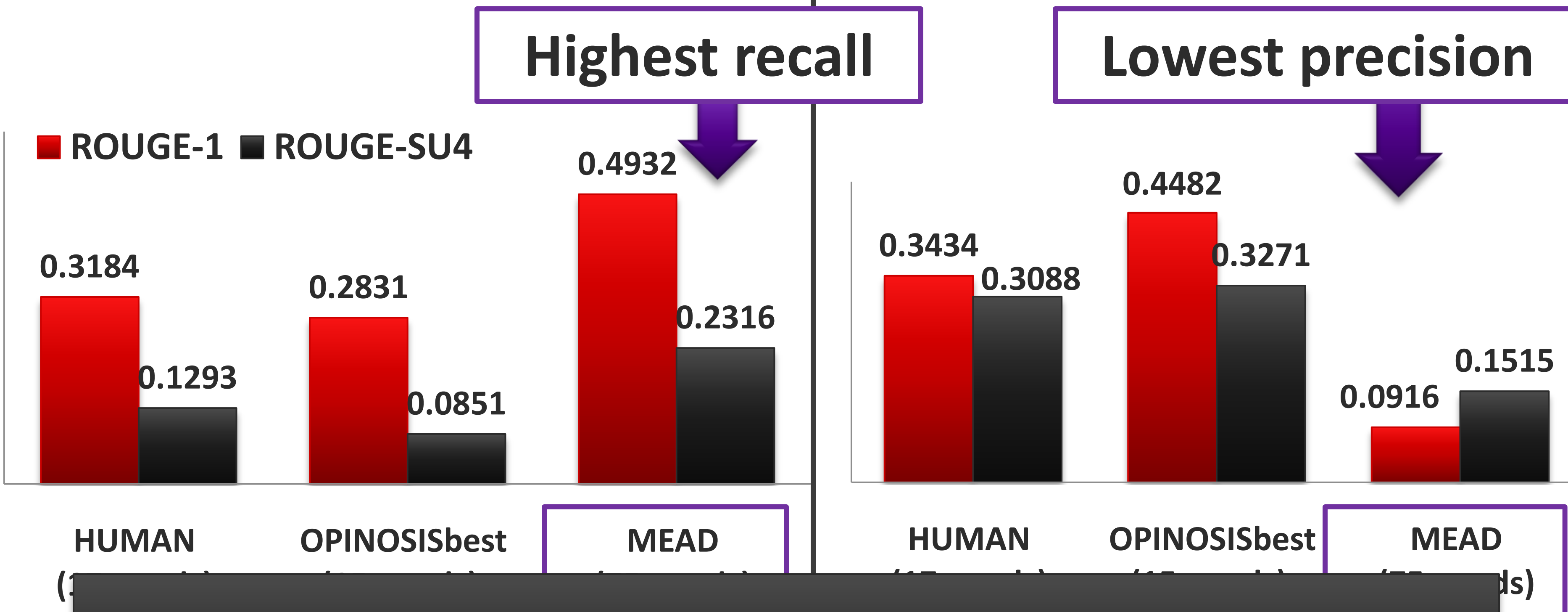


ROUGE Recall



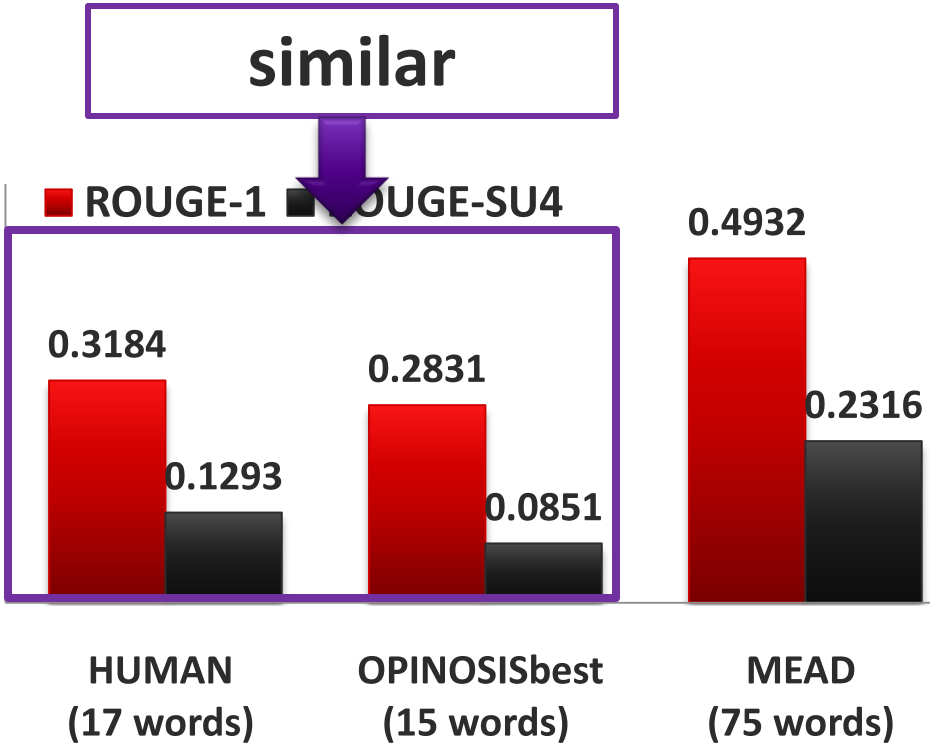
ROUGE Precision

Human vs. Opinosis vs. MEAD

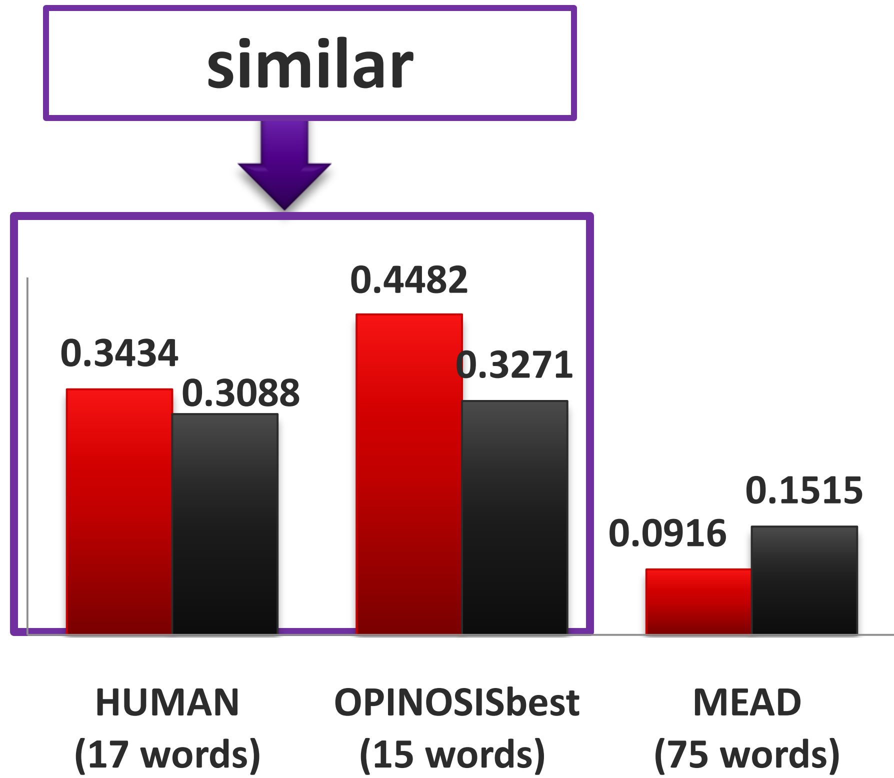


Overall: Baseline does not do well in generating concise summaries.

Human vs. Opinosis

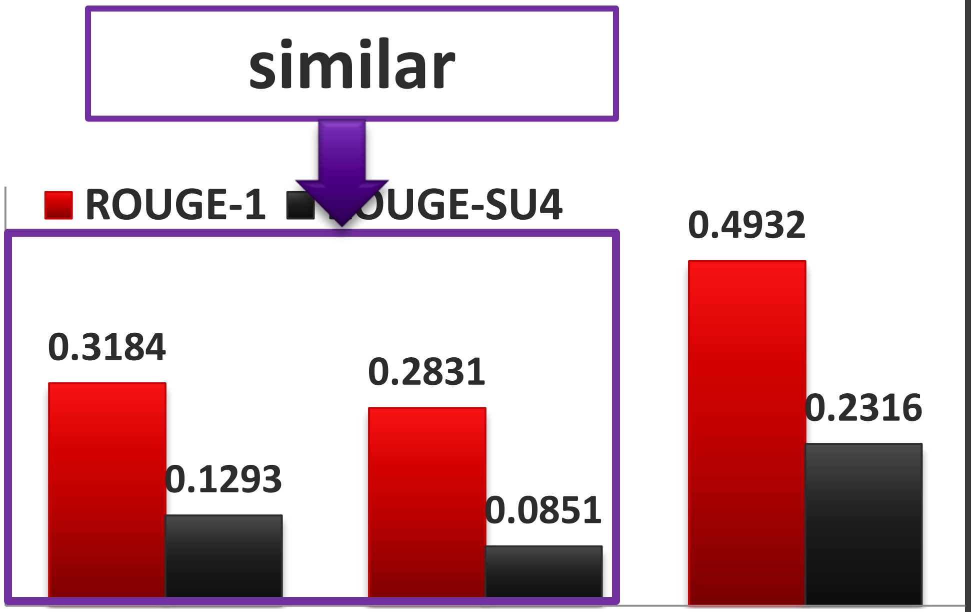


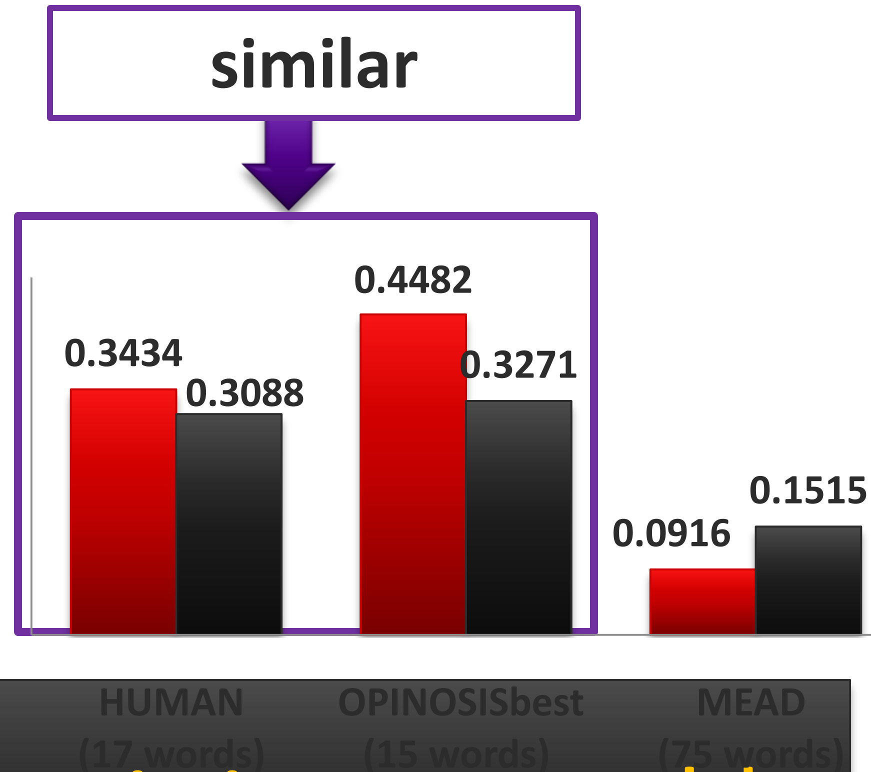
ROUGE Recall



ROUGE Precision

Human vs. Opinosis



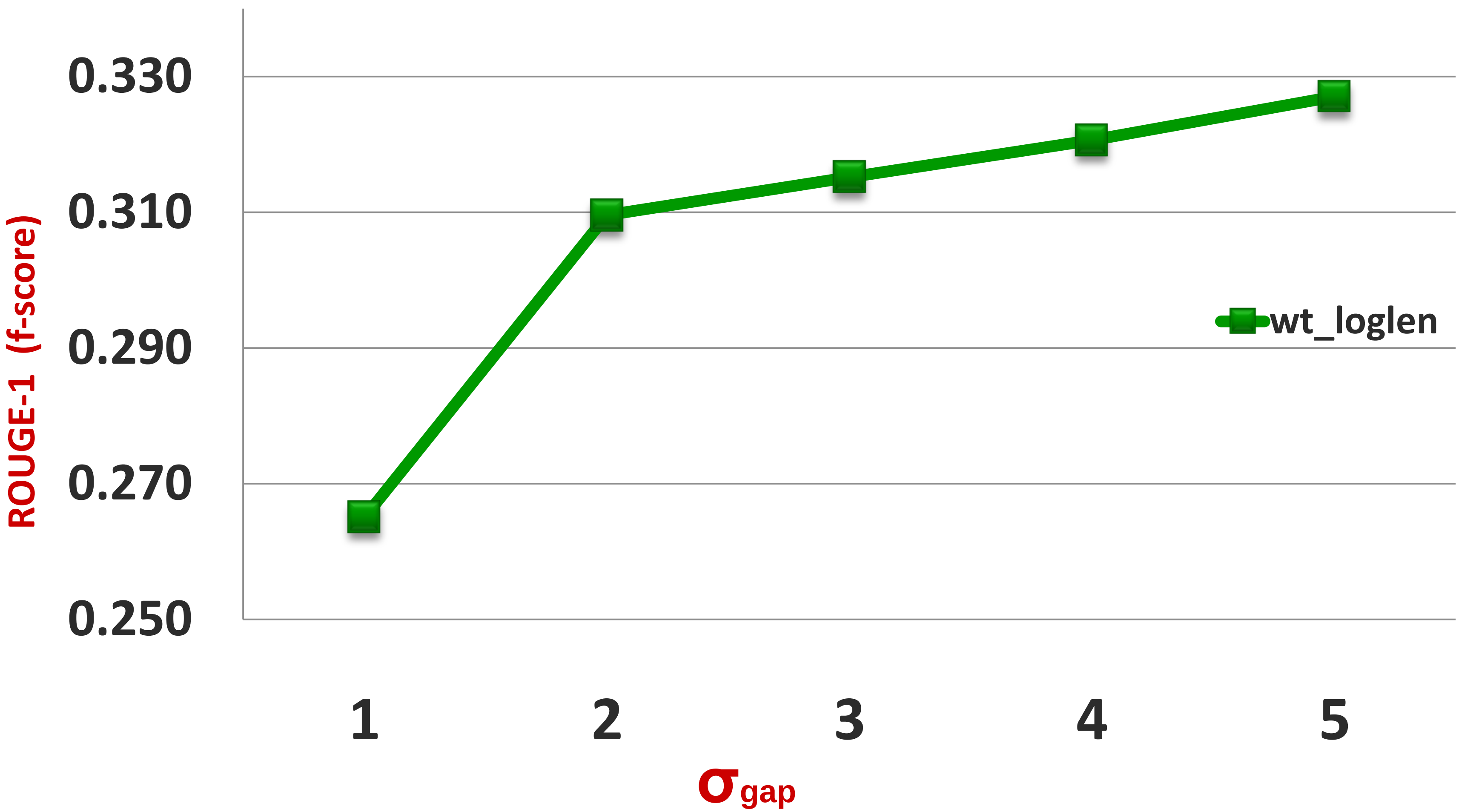


Performance of Opinosis is reasonable

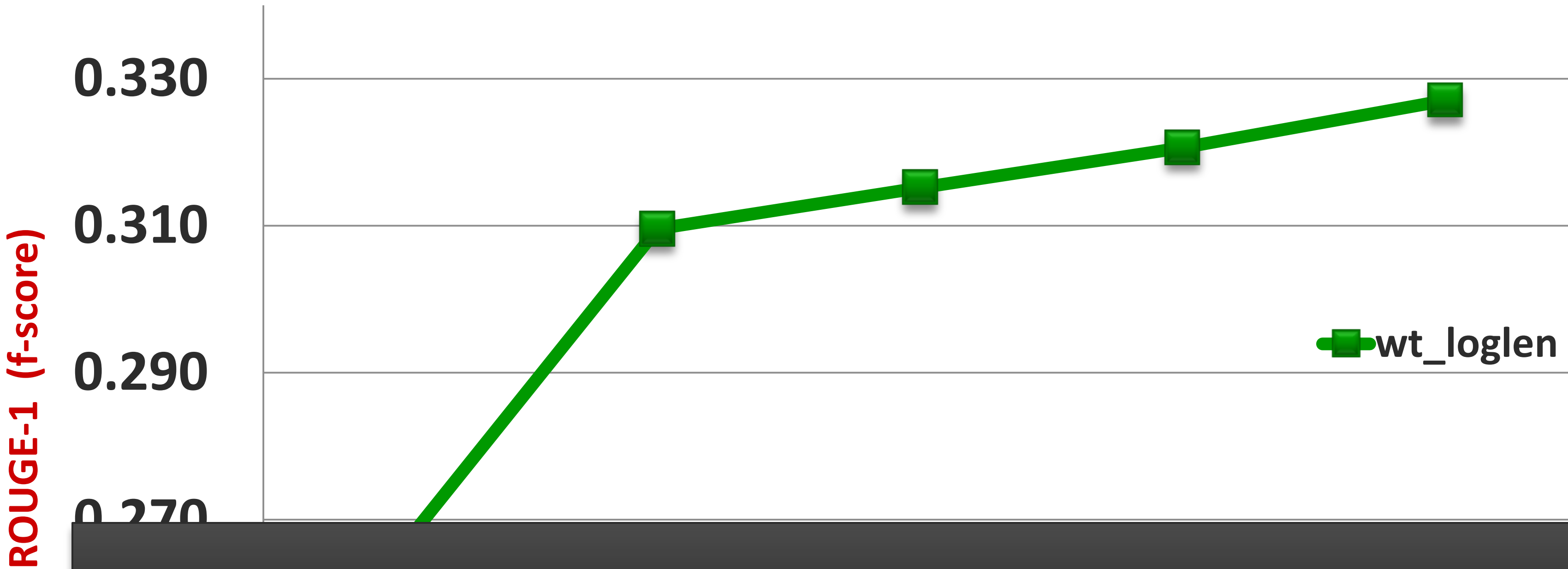
→ similar to Human performance

Effect of Gap Threshold (σ_{gap})

Effect of Gap Threshold (σ_{gap})

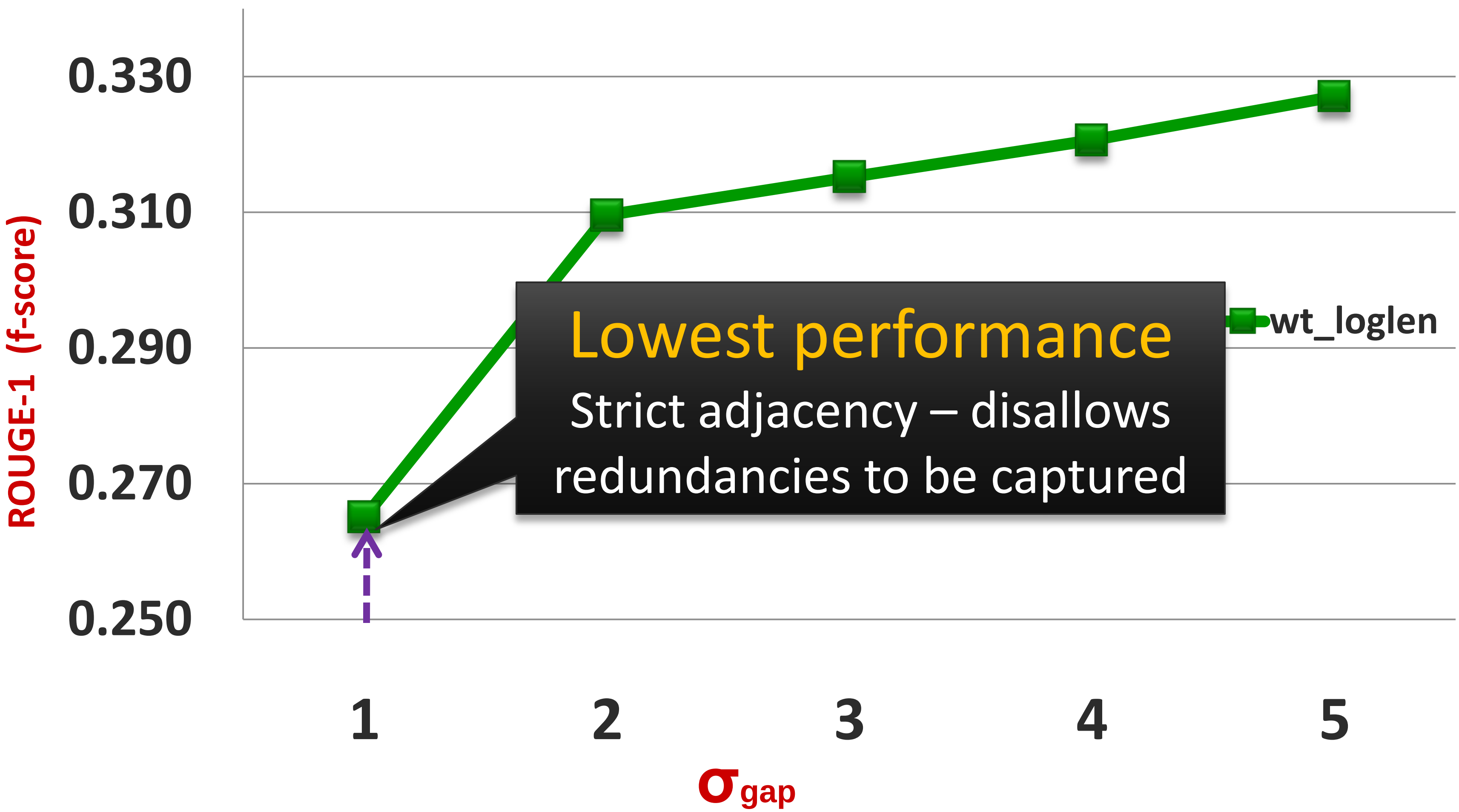


Effect of Gap Threshold (σ_{gap})

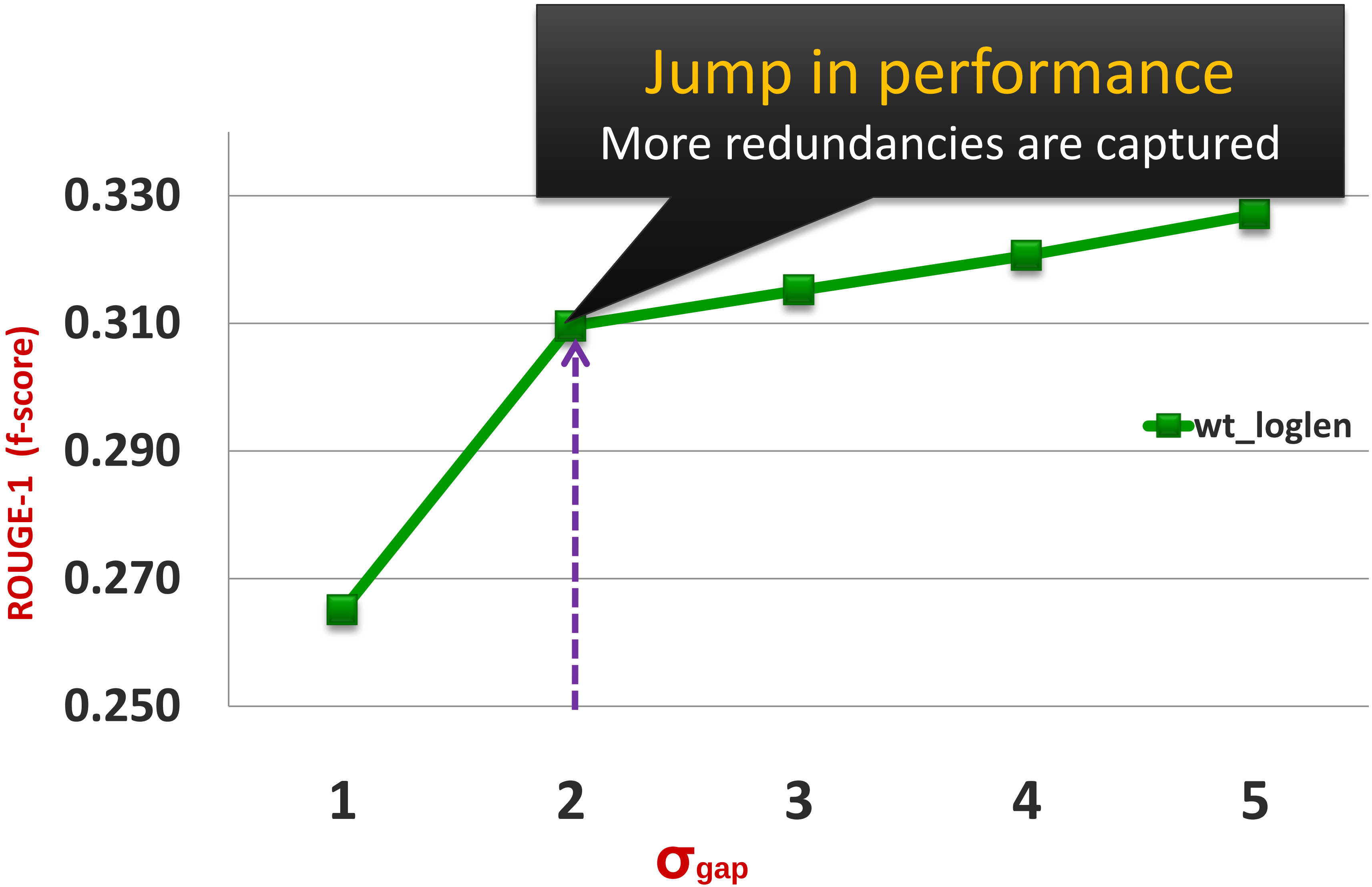


small σ_{gap} → words in summary close together in original text

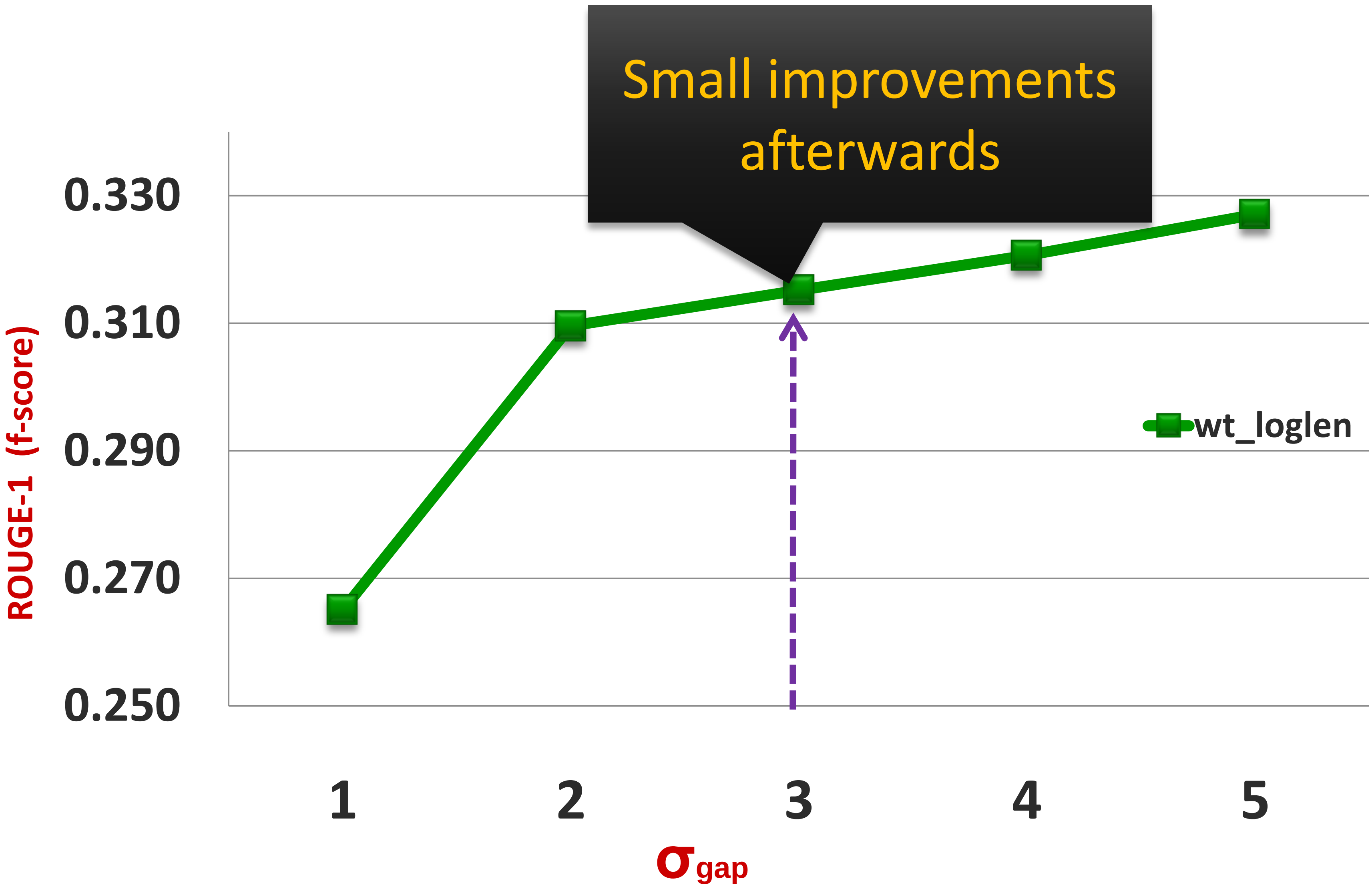
Effect of Gap Threshold (σ_{gap})



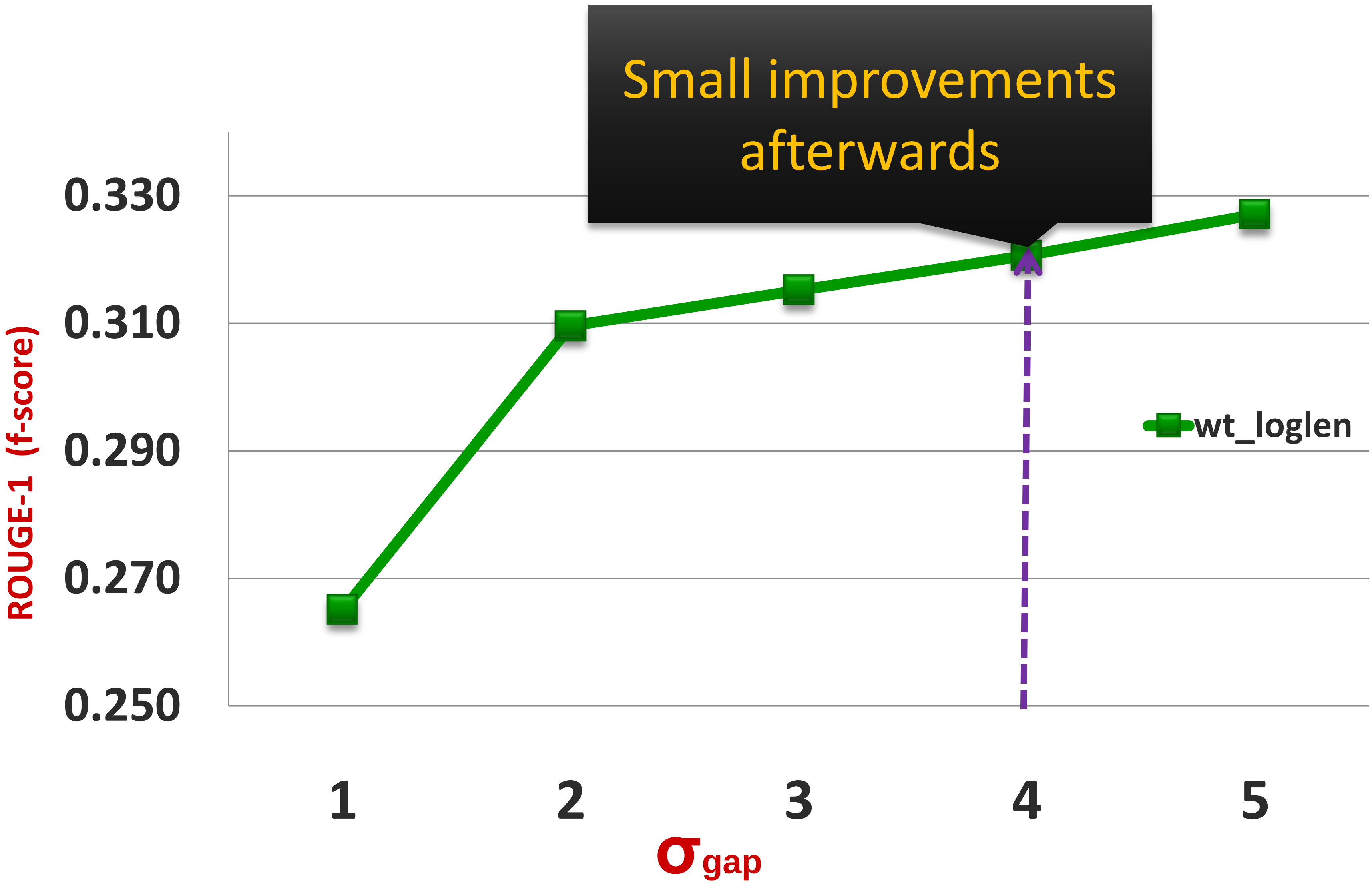
Effect of Gap Threshold (σ_{gap})



Effect of Gap Threshold (σ_{gap})



Effect of Gap Threshold (σ_{gap})



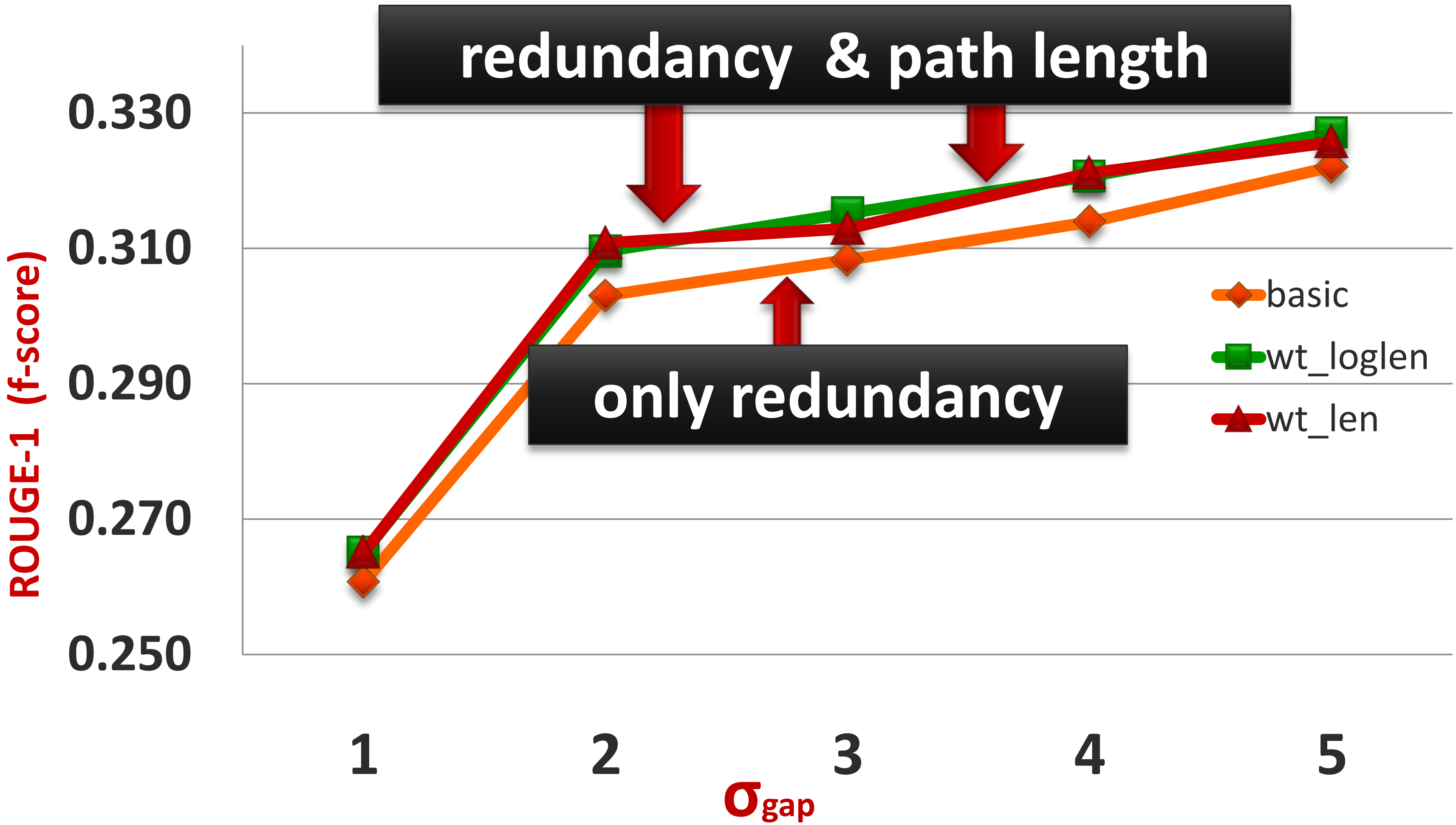
Effect of Gap Threshold (σ_{gap})

σ_{gap} too large: ill formed sentences

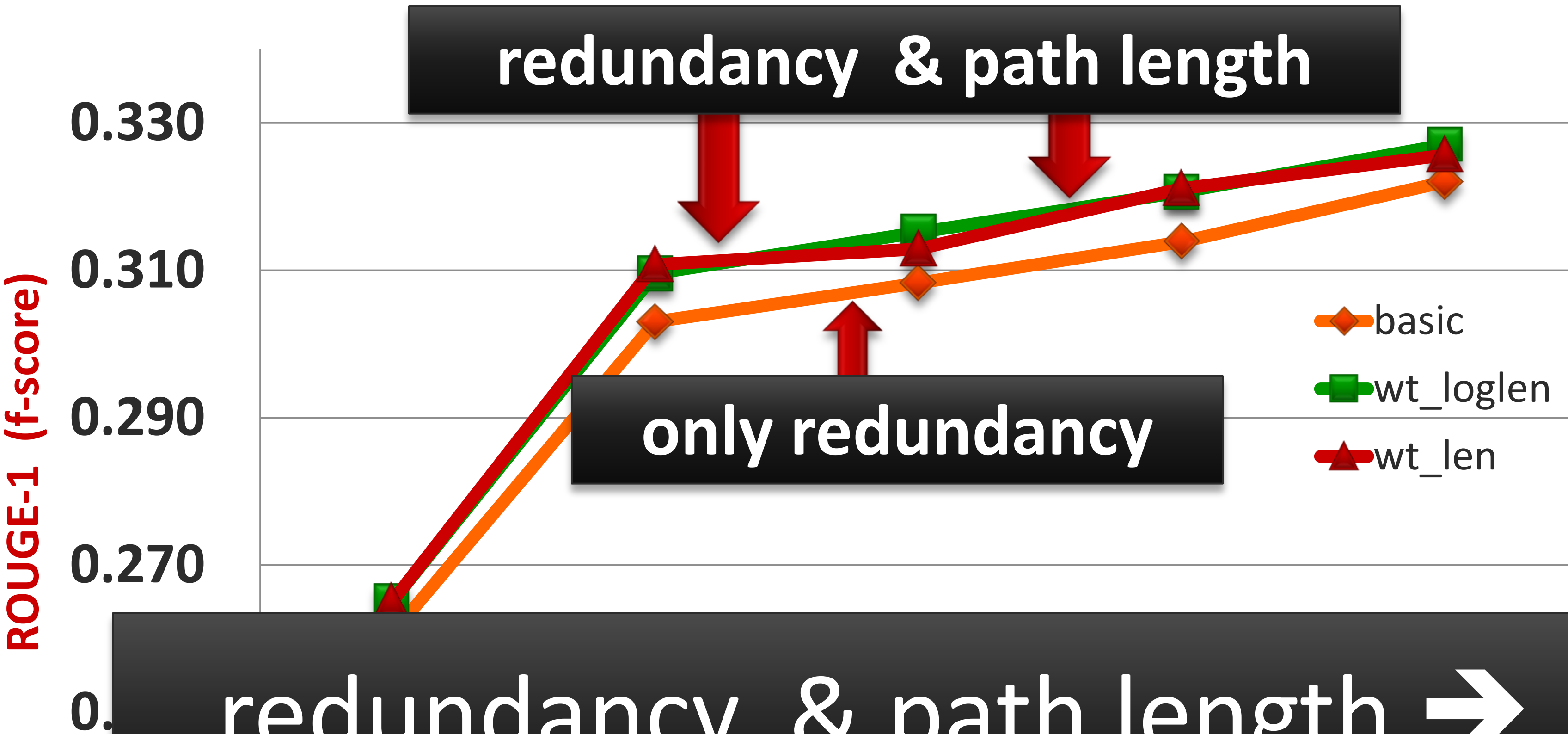
Set σ_{gap} to low value

Compare: Scoring Functions

Compare: Scoring Functions

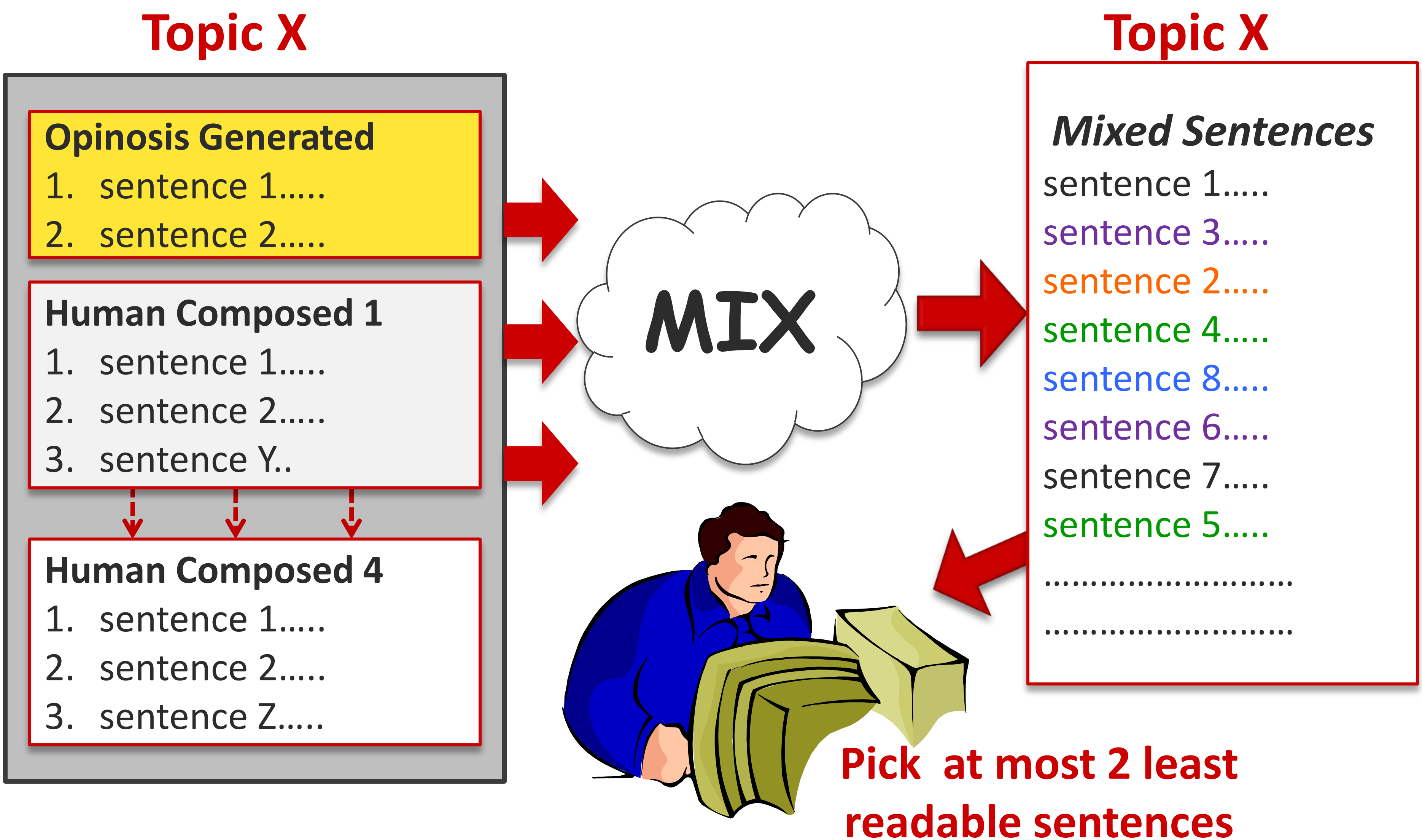


Compare: Scoring Functions



redundancy & path length →
summaries with better coverage

Readability Test



Readability Test

- **Assessor often picks:**
  - **Opinosis** sentences - Opinosis summaries have readability issues
  - **Non-Opinosis sentences or makes no picks** - Opinosis summaries similar to human summaries

Our Readability Test

- **Assessor picked:**
  - **34/102** Opinosis generated sentences as least readable

Our Readability Test

- **Assessor picked:**
  - **34/102** Opinosis generated sentences as least readable

> 60% of Opinosis sentences are not very 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)

Dataset and Demo Software is available

http://timan.cs.uiuc.edu/downloads.html

References

[Barzilay and Lee2003] Barzilay, Regina and Lillian Lee. 2003. Learning to paraphrase: an unsupervised approach using multiple-sequence alignment. In NAACL ’03: Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, pages 16–23, Morristown, NJ, USA.

[DeJong1982] DeJong, Gerald F. 1982. An overview of the FRUMP system. In Lehnert, Wendy G. and Martin H. Ringle, editors, Strategies for Natural Language Processing, pages 149–176. Lawrence Erlbaum, Hillsdale, NJ.

[Erkan and Radev2004] Erkan, G¨unes and Dragomir R. Radev. 2004. Lexrank: graph-based lexical centrality as salience in text summarization. J. Artif. Int. Res.,22(1):457–479.

[Finley and Harabagiu2002] Finley, Sanda Harabagiu and Sanda M. Harabagiu. 2002. Generating single and multi-document summaries with gistexter. In Proceedings of the workshop on automatic summarization, pages 30–38.

[Hu and Liu2004] Hu, Minqing and Bing Liu. 2004. Mining and summarizing customer reviews. In KDD, pages 168–177.

[Jing and McKeown2000] Jing, Hongyan and Kathleen R. McKeown. 2000. Cut and paste based text summarization. In Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference, pages 178–185, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

[Lerman et al.2009] Lerman, Kevin, Sasha Blair-Goldensohn, and Ryan Mcdonald. 2009. Sentiment summarization: Evaluating and learning user preferences. In 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL-09).

[Lin and Hovy2003] Lin, Chin-Yew and Eduard Hovy. 2003. Automatic evaluation of summaries using ngram co-occurrence statistics. In Proc. HLT-NAACL, page 8 pages.

[LIN2004a] LIN, Chin-Yew. 2004a. Looking for a few good metrics : Rouge and its evaluation. proc. of the 4th NTCIR Workshops, 2004.

[Lin2004b] Lin, Chin-Yew. 2004b. Rouge: a package for automatic evaluation of summaries. In Proceedings of theWorkshop on Text Summarization Branches Out (WAS 2004), Barcelona, Spain.

[Lu et al.2009] Lu, Yue, ChengXiang Zhai, and Neel Sundaresan. 2009. Rated aspect summarization of short comments. In 18th International World WideWeb Conference (WWW2009), April.

..References

[Mihalcea and Tarau2004] Mihalcea, R. and P. Tarau. 2004. TextRank: Bringing order into texts. In Proceedings of EMNLP-04and the 2004 Conference on Empirical Methods in Natural Language Processing, July.

[Pang and Lee2004] Pang, Bo and Lillian Lee. 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In Proceedings of the ACL, pages 271–278.

[Pang et al.2002] Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 79–86.

[Radev and McKeown1998] Radev, DR and K. McKeown. 1998. Generating natural language summaries from multiple on-line sources. Computational Linguistics, 24(3):469–500.

[Radev et al.2000] Radev, Dragomir, Hongyan Jing, and Malgorzata Budzikowska. 2000. Centroid-based summarization of multiple documents: Sentence extraction, utility-based evaluation, and user studies. In ANLP/NAACL Workshop on Summarization, pages 21–29.

[Radev et al.2002] Radev, Dragomir R., Eduard Hovy, and Kathleen McKeown. 2002. Introduction to the special issue on summarization.

[Saggion and Lapalme2002] Saggion, Horacio and Guy Lapalme. 2002. Generating indicative-informative summaries with sumum. Computational Linguistics, 28(4):497–526.

[Snyder and Barzilay2007] Snyder, Benjamin and Regina Barzilay. 2007. Multiple aspect ranking using the good grief algorithm. In In Proceedings of the Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics (HLT-NAACL), pages 300–307.

[Titov and Mcdonald2008] Titov, Ivan and Ryan Mcdonald. 2008. A joint model of text and aspect ratings for sentiment summarization. In Proceedings of ACL-08: HLT, pages 308–316, Columbus, Ohio, June. Association for Computational Linguistics.

Thank You...

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

TIMan