Using Tweets to Help Sentence Compression for News Highlights Generation

Zhongyu Wei[1], Yang Liu[1], Chen Li[1], Wei Gao[2]

[1]Computer Science Department, The University of Texas at Dallas, Texas, USA
[2]Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar

July 27th 2015 Beijing, China
The 53rd Annual Meetings of the Association for Computational Linguistics

NASA probe set to make history at Pluto

By Amanda Barnett, CNN
Updated 2:48 PM ET, Fri July 10, 2015 | Video Source: CNN/NASA

Story highlights

- The New Horizons spacecraft is closing in on Pluto
- The probe is first to do a flyby of the dwarf planet
- The mission will complete a NASA tour of the classical solar system

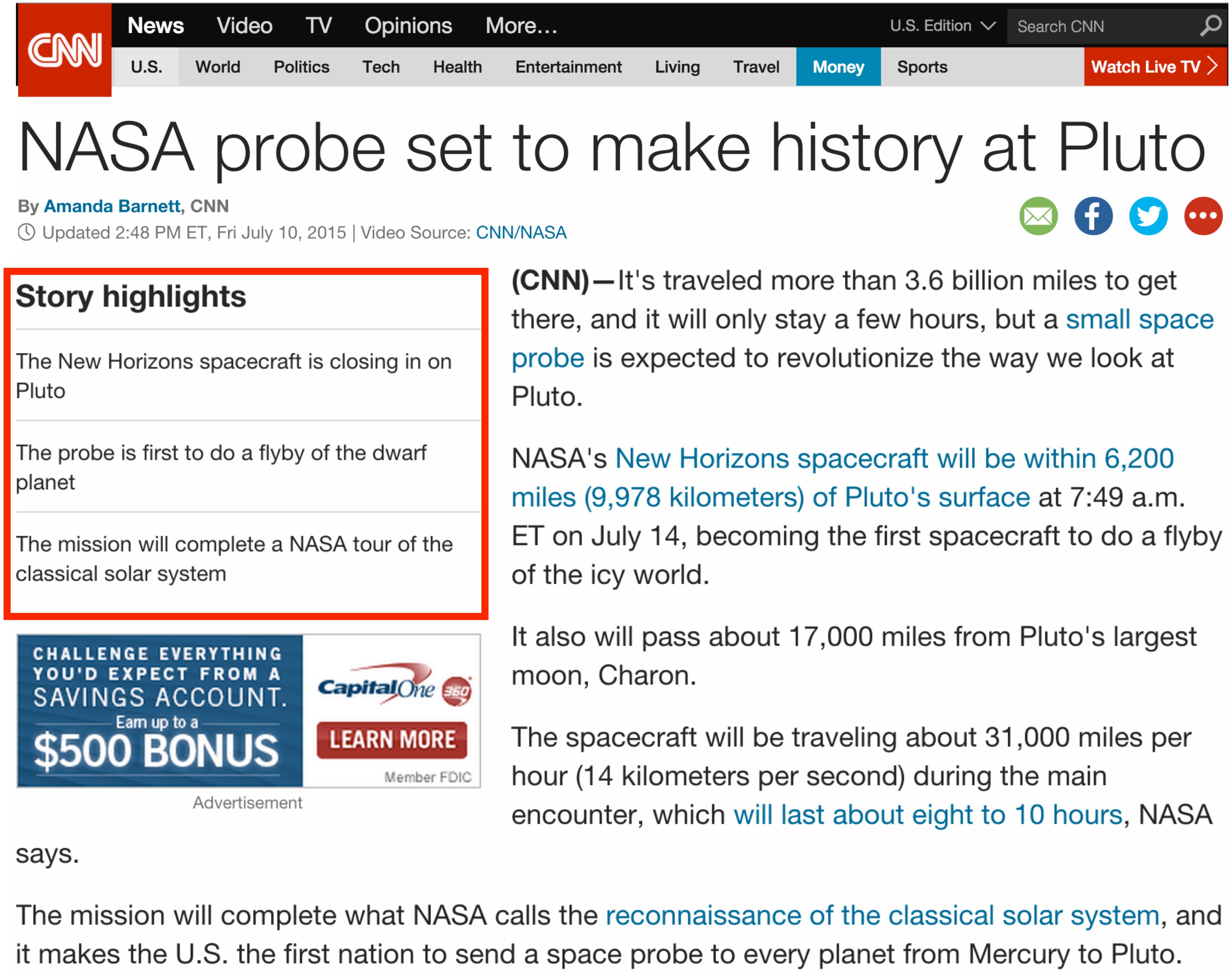
(CNN)—It's traveled more than 3.6 billion miles to get there, and it will only stay a few hours, but a small space probe is expected to revolutionize the way we look at Pluto.

NASA's New Horizons spacecraft will be within 6,200 miles (9,978 kilometers) of Pluto's surface at 7:49 a.m. ET on July 14, becoming the first spacecraft to do a flyby of the icy world.

It also will pass about 17,000 miles from Pluto's largest moon, Charon.

The spacecraft will be traveling about 31,000 miles per hour (14 kilometers per second) during the main encounter, which will last about eight to 10 hours, NASA says.

The mission will complete what NASA calls the reconnaissance of the classical solar system, and it makes the U.S. the first nation to send a space probe to every planet from Mercury to Pluto.



NASA probe set to make history at Pluto

By Amanda Barnett, CNN
Updated 2:48 PM ET, Fri July 10, 2015 | Video Source: CNN/NASA

Story highlights

- The New Horizons spacecraft is closing in on Pluto
- The probe is first to do a flyby of the dwarf planet
- The mission will complete a NASA tour of the classical solar system

(CNN)—It's traveled more than 3.6 billion miles to get there, and it will only stay a few hours, but a small space probe is expected to revolutionize the way we look at Pluto.

NASA's New Horizons spacecraft will be within 6,200 miles (9,978 kilometers) of Pluto's surface at 7:49 a.m. ET on July 14, becoming the first spacecraft to do a flyby of the icy world.

It also will pass about 17,000 miles from Pluto's largest moon, Charon.

The spacecraft will be traveling about 31,000 miles per hour (14 kilometers per second) during the main encounter, which will last about eight to 10 hours, NASA says.

The mission will complete what NASA calls the reconnaissance of the classical solar system, and it makes the U.S. the first nation to send a space probe to every planet from Mercury to Pluto.

Highlight sentence is much shorter than the original sentence. Extractive summarization is not enough!

Sentence Compression for Summarization

- Sentence compression aims to retain the most valuable information of an original sentence.

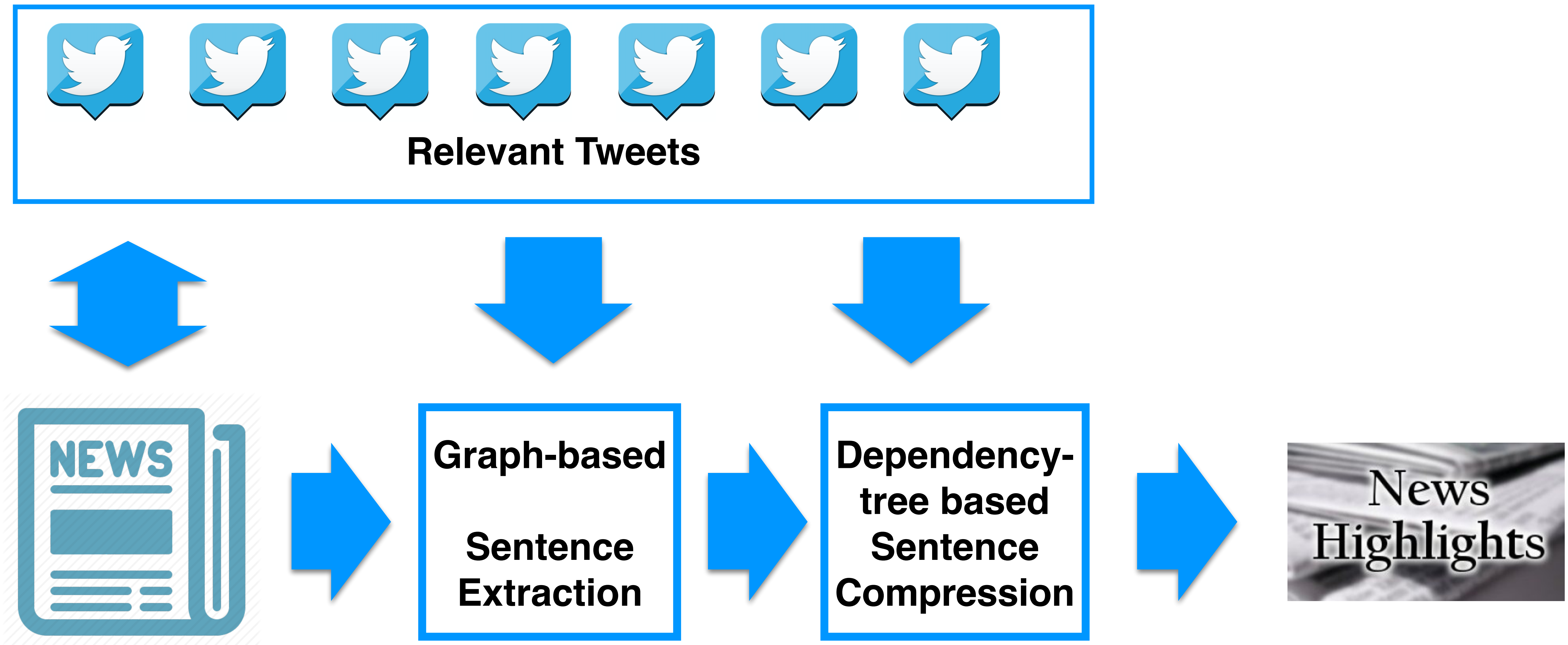
Sentence Compression for Summarization

- Sentence compression aims to retain the most valuable information of an original sentence.
- Pipeline system combining sentence extraction and compression
  - Generic sentence compression module does not consider summarization task

Sentence Compression for Summarization

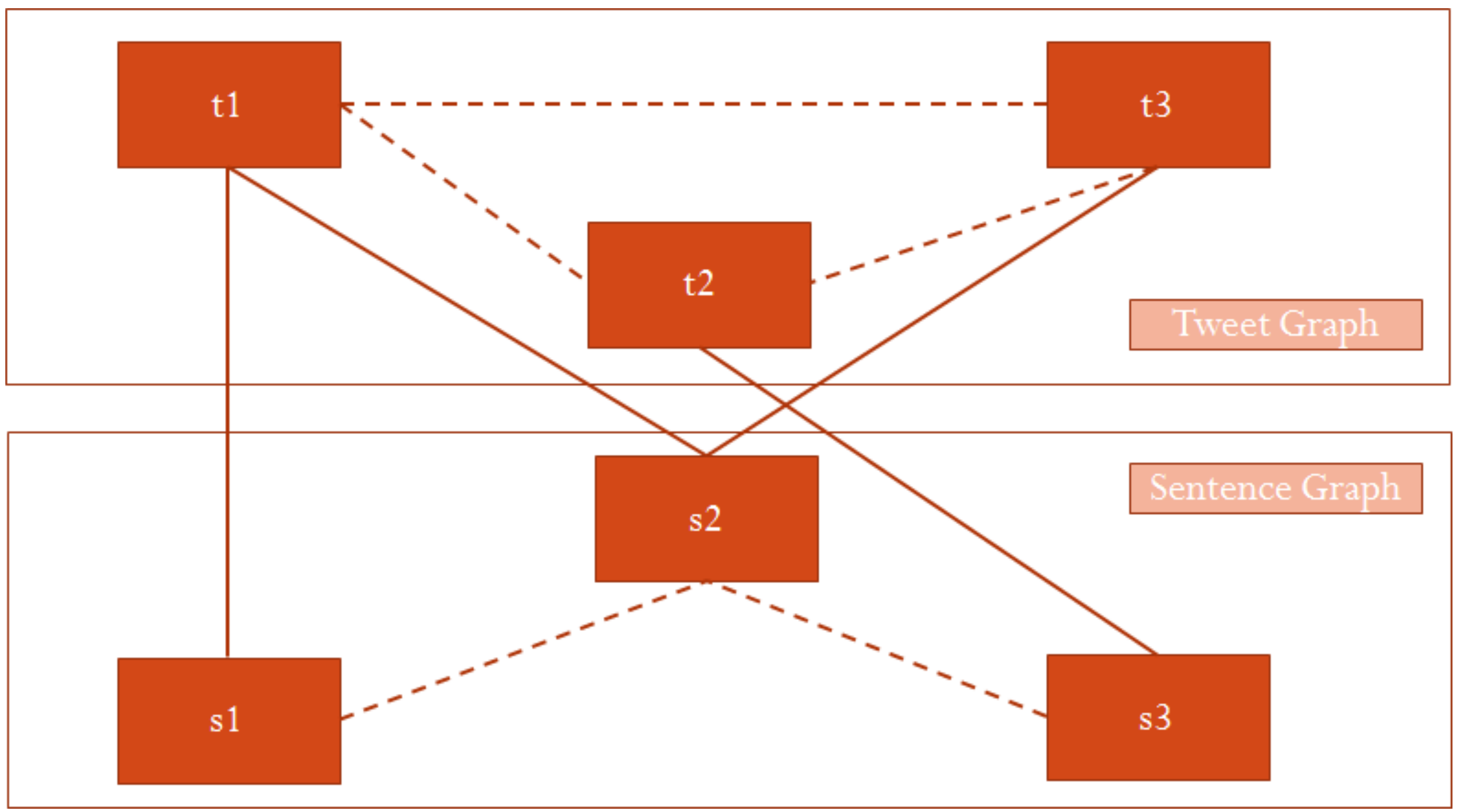
- Sentence compression aims to retain the most valuable information of an original sentence.
- Pipeline system combining sentence extraction and compression
  - Generic sentence compression module does not consider summarization task
- Summary guided compression method (Li et. al. EMNLP'13)
  - It relies on manually generated corpus

Using Tweets for Compressive Summarization



Tweets Involved Sentence Extraction

- Heterogeneous Graph Random Walk (HGRW) (Wei and Gao, SigIR'15)



$$sim(x,y) = \left\{ \begin{array}{ll} \varepsilon * consine(x,y) & if \;\; x.t \neq y.t \\ (1-\varepsilon) * consine(x,y) & otherwise \end{array} \right\}$$

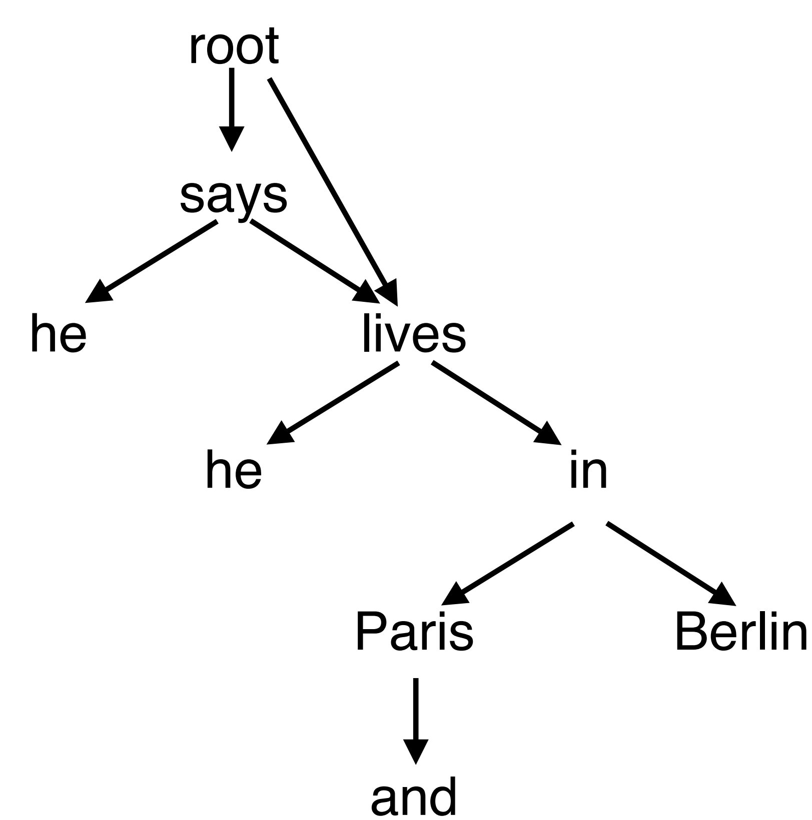
Sentence Compression

Given a sentence *s*, consisting of words w_1, w_2, …, w_m, identify a subset.

- Dependency-tree-based

  He says he lives in Paris and Berlin

  1. Dependency graph generation
  2. Tree structure pruning



Sentence Compression

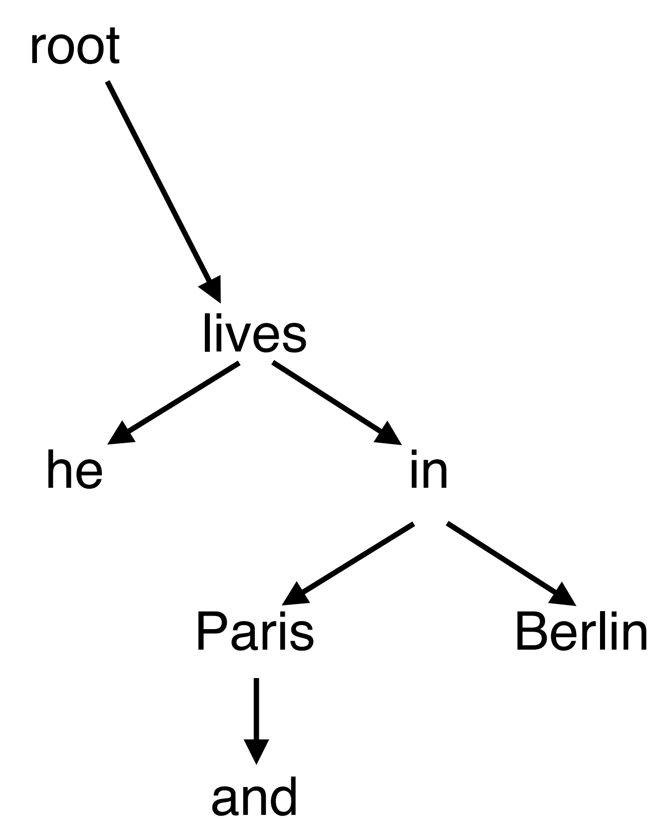
Given a sentence *s*, consisting of words w_1, w_2, …, w_m, identify a subset.

- Dependency-tree-based

  He says he lives in Paris and Berlin

  1. Dependency graph generation
  2. Tree structure pruning

  he lives in Paris and Berlin



Dependency Tree Based Sentence Compression

(Filippova and Strube, INLGC'08)

- Find a subtree with highest objective score:
  - e is an edge, x_e is binary variable
  - $\boldsymbol{w_{info}(e)}$ is the informativeness
  - $\boldsymbol{w_{syn}(e)}$ is the syntactic importance

$$f(X) = \sum_{e \in E} x_e \times w_{info}(e) \times w_{syn}(e)$$

- ILP is used for solving the problem with some constraints

Baseline Weighting Approach

head (h) $\xrightarrow{\text{label (l)}}$ node (n)

- Informativeness
  - $P_{summary}(n)$ is the unigram probability from human generated summaries
  - $P_{article}(n)$ is the unigram probability from original articles

$$w_{info}(e) = \frac{P_{summary}(n)}{P_{article}(n)}$$

- Syntactic importance
  - $P(l|h)$ is the conditional probability of label *l* given head *h*.

$$w_{syn}(e) = P(l \mid h)$$

Leverage Tweets for Edge Weighting

$$w_{\inf o}(e) = (1-\alpha)\cdot w_{\inf o}^{N}(e) + \boxed{\alpha \cdot w_{\inf o}^{T}(e)}$$

$$w_{syn}(e) = (1-\beta)\cdot w_{syn}^{N}(e) + \boxed{\beta \cdot w_{syn}^{T}(e)}$$

- Informativeness
  - $P_{relevent}(n)$ is from relevant tweets to the news
  - $P_{background}(n)$ is from a background tweet corpus

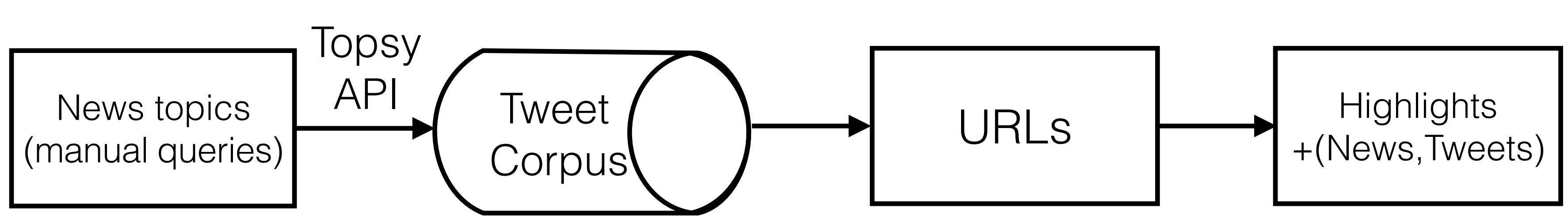
$$w_{\inf o}^{T}(e) = \frac{P_{relevantT}(n)}{P_{backgroundT}(n)}$$

- Syntactic importance
  - NT(h,n) is the number of tweets contain both h and n
  - NT is the total number of relevant tweets

$$w_{syn}^{T}(e) = \frac{NT(h,n)}{NT}$$

Dataset Collection

- Tweets gathering using Topsy API for 17 topics
- News articles from CNN.com and USAToday.com
- Precise version: link news to tweets with its URLs
- Broad version: link news to tweets related to same topic



Dataset Collection (cont.’)

- Distribution of news, highlights, and tweets per topic
  - 121 news articles, 455 highlight sentences
  - 78,419 linked tweets and 6,890,987 retrieved tweets

Event	Doc #	HLight #	Linked Tweet #	Retrieved Tweet #	Event	Doc #	HLight #	Linked Tweet #	Retrieved Tweet #
Aurora shooting	14	54	12,463	588,140	African runner murder	8	29	9,461	303,535
Boston bombing	38	147	21,683	1,650,650	Syria chemical weapons use	1	4	331	11,850
Connecticut shooting	13	47	3,021	213,864	US military in Syria	2	7	719	619,22
Edward Snowden	5	17	1,955	379,349	DPRK Nuclear Test	2	8	3,329	103,964
Egypt balloon crash	3	12	836	36,261	Asiana Airlines Flight 214	11	42	8,353	351,412
Hurricane Sandy	4	15	607	189,082	Moore Tornado	5	19	1,259	1,154,656
Russian meteor	3	11	6,841	239,281	Chinese Computer Attacks	2	8	507	28,988
US Flu Season	7	23	6,304	1,042,169	Williams Olefins Explosion	1	4	268	14,196
Super Bowl blackout	2	8	482	214,775	Total	121	455	78,419	6,890,987

Experiment Setup

- Sentence extraction
  - LexRank
  - HGRW (precise version relevant tweet dataset)
- Sentence compression
  - Baseline (Filippova and Strube, INLGC'08)
    - using NYT news corpus for edge weighting
  - Tweets related edge weighting
    - Informativeness (precise version relevant tweet dataset)
    - Syntactic (broad version relevant tweet dataset)
- Evaluation
  - ROUGE-1 score

Overall Experiment Result

Method	ROUGE-1			Compr. Rate(%)
	F(%)	P(%)	R(%)	
LexRank	26.1	19.9	39.1	100
LexRank+SC	25.2	22.4	29.6	63.0
LexRank+SC+w^T_{info}	25.7	22.8	30.1	62.0
LexRank+SC+w^T_{syn}	26.2	23.5	30.4	63.7
LexRank+SC+both	**27.5**	**25.0**	31.4	61.5
HGRW	28.1	22.6	**39.5**	100
HGRW+SC	26.4	24.9	29.5	66.1
HGRW+SC+w^T_{info}	27.5	25.7	30.8	65.4
HGRW+SC+w^T_{syn}	27.0	25.3	30.2	66.7
HGRW+SC+both	**28.4**	**26.9**	31.2	64.8

+SC: with baseline sentence compression
+w^T_{***}: with tweets for edge weighting for informativeness or syntactic importance
+both: use both tweets related factors for edge weighting

Overall Experiment Result

Method	ROUGE-1			Compr. Rate(%)
	F(%)	P(%)	R(%)	
LexRank	26.1	19.9	39.1	100
LexRank+SC	25.2	22.4	29.6	63.0
LexRank+SC+w^T_{info}	25.7	22.8	30.1	62.0
LexRank+SC+w^T_{syn}	26.2	23.5	30.4	63.7
LexRank+SC+both	**27.5**	**25.0**	31.4	61.5
HGRW	28.1	22.6	**39.5**	100
HGRW+SC	26.4	24.9	29.5	66.1
HGRW+SC+w^T_{info}	27.5	25.7	30.8	65.4
HGRW+SC+w^T_{syn}	27.0	25.3	30.2	66.7
HGRW+SC+both	**28.4**	**26.9**	31.2	64.8

+SC: with baseline sentence compression

+w^T_{***}: with tweets for edge weighting for informativeness or syntactic importance

+both: use both tweets related factors for edge weighting

Overall Experiment Result

Method	ROUGE-1			Compr. Rate(%)
	F(%)	P(%)	R(%)	
LexRank	26.1	19.9	39.1	100
LexRank+SC	25.2	22.4	29.6	63.0
LexRank+SC+w^T_{info}	25.7	22.8	30.1	62.0
LexRank+SC+w^T_{syn}	26.2	23.5	30.4	63.7
LexRank+SC+both	**27.5**	**25.0**	31.4	61.5
HGRW	28.1	22.6	**39.5**	100
HGRW+SC	26.4	24.9	29.5	66.1
HGRW+SC+w^T_{info}	27.5	25.7	30.8	65.4
HGRW+SC+w^T_{syn}	27.0	25.3	30.2	66.7
HGRW+SC+both	**28.4**	**26.9**	31.2	64.8

+SC: with baseline sentence compression
+w^T_{***}: with tweets for edge weighting for informativeness or syntactic importance
+both: use both tweets related factors for edge weighting

Overall Experiment Result

Method	ROUGE-1			Compr. Rate(%)
	F(%)	P(%)	R(%)	
LexRank	26.1	19.9	39.1	100
LexRank+SC	25.2	22.4	29.6	63.0
LexRank+SC+w^T_{info}	25.7	22.8	30.1	62.0
LexRank+SC+w^T_{syn}	26.2	23.5	30.4	63.7
LexRank+SC+both	<u>**27.5**</u>	<u>**25.0**</u>	31.4	61.5
HGRW	28.1	22.6	**39.5**	100
HGRW+SC	26.4	24.9	29.5	66.1
HGRW+SC+w^T_{info}	27.5	25.7	30.8	65.4
HGRW+SC+w^T_{syn}	<u>27.0</u>	<u>25.3</u>	30.2	66.7
HGRW+SC+both	**28.4**	**26.9**	31.2	64.8

+SC: with baseline sentence compression

+$w^{T_{***}}$: with tweets for edge weighting for informativeness or syntactic importance

+both: use both tweets related factors for edge weighting

Overall Experiment Result

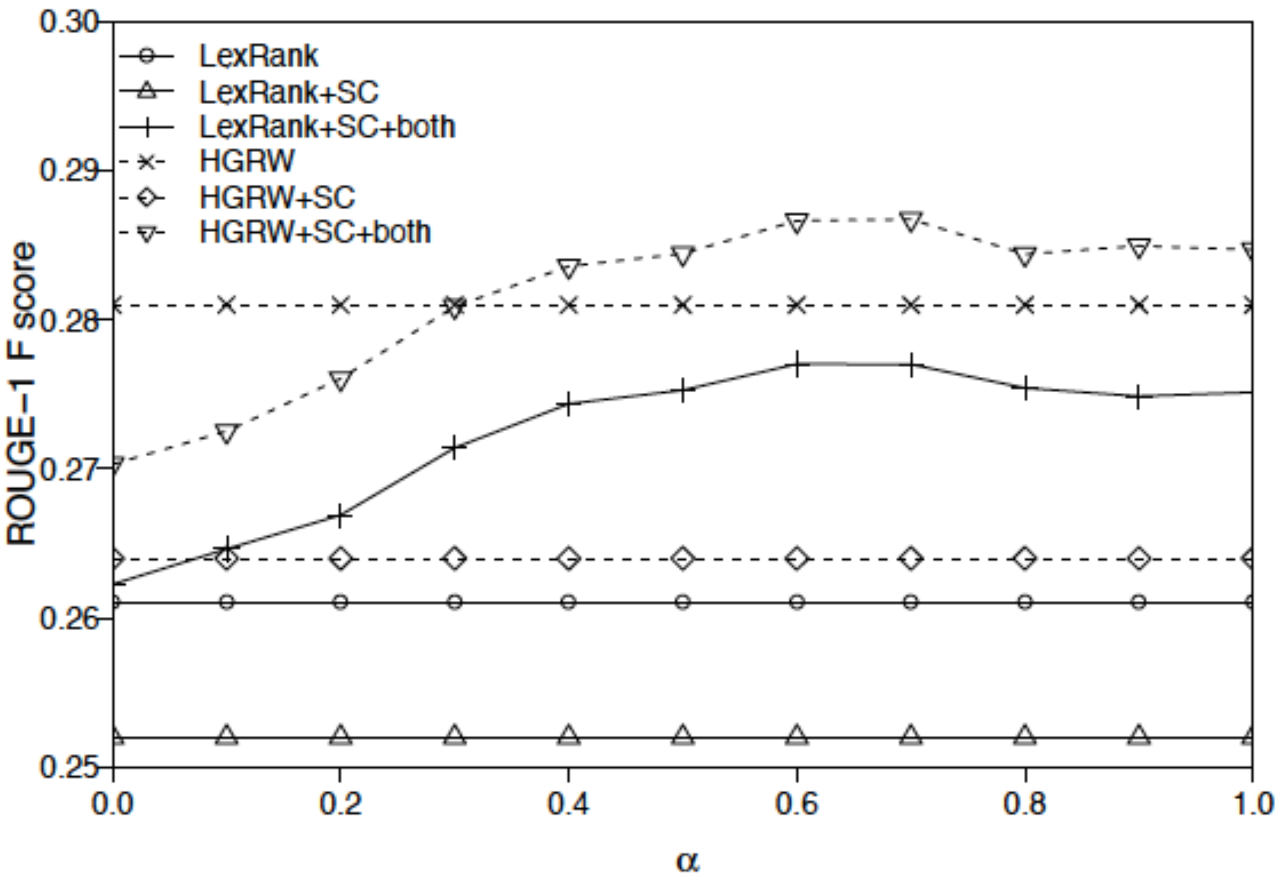
Method	ROUGE-1			Compr. Rate(%)
	F(%)	P(%)	R(%)	
LexRank	26.1	19.9	39.1	100
LexRank+SC	25.2	22.4	29.6	63.0
LexRank+SC+w^T_{info}	25.7	22.8	30.1	62.0
LexRank+SC+w^T_{syn}	26.2	23.5	30.4	63.7
LexRank+SC+both	**27.5**	**25.0**	31.4	61.5
HGRW	28.1	22.6	**39.5**	100
HGRW+SC	26.4	24.9	29.5	66.1
HGRW+SC+w^T_{info}	27.5	25.7	30.8	65.4
HGRW+SC+w^T_{syn}	27.0	25.3	30.2	66.7
HGRW+SC+both	**28.4**	**26.9**	31.2	64.8

+SC: with baseline sentence compression
+w^T_{***}: with tweets for edge weighting for informativeness or syntactic importance
+both: use both tweets related factors for edge weighting

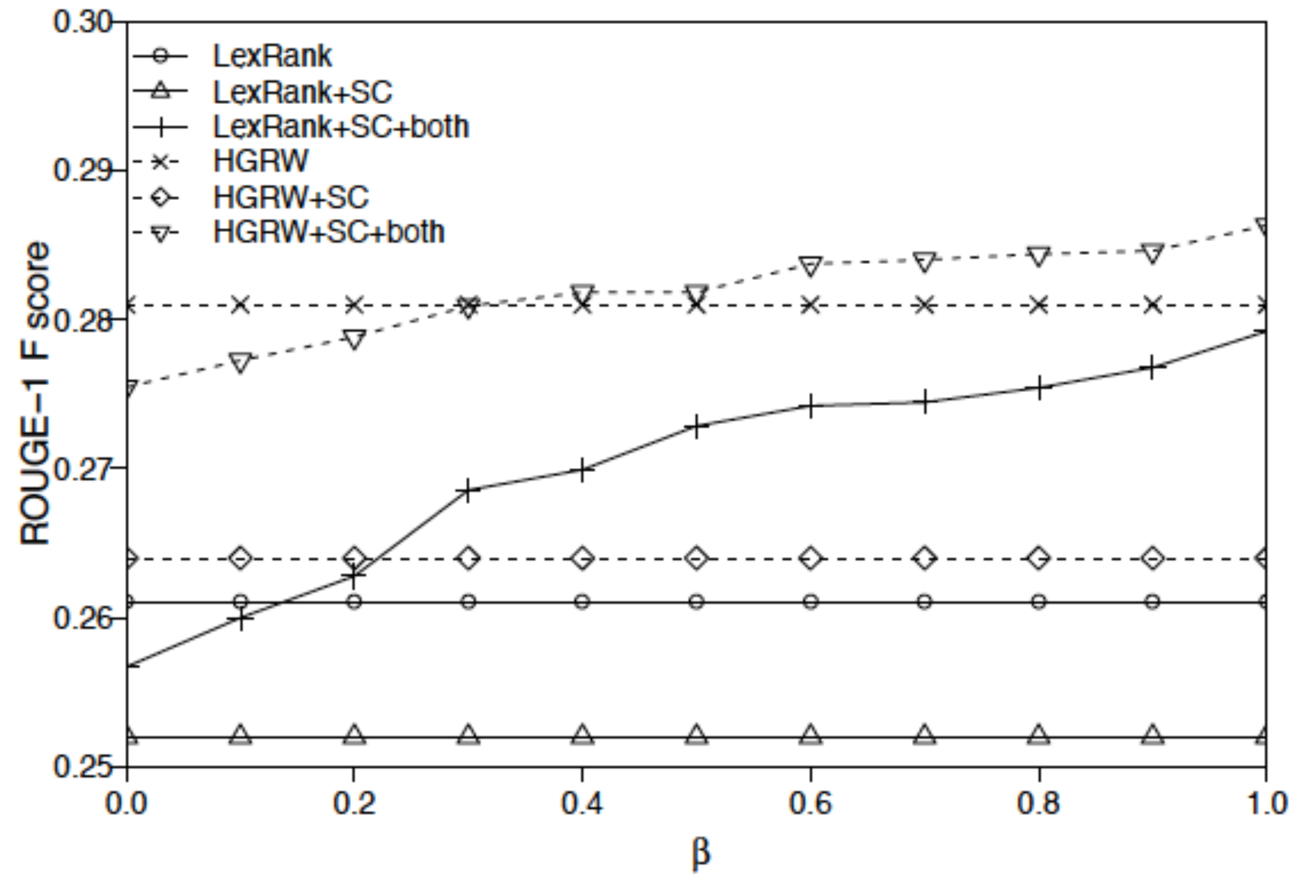
Example

Method	**Result**
LexRank	Boston bombing suspect Tamerlan Tsarnaev, killed in a shootout with police days after the blast, has been buried at an undisclosed location, police in Worcester, Mass., said.
LexRank+SC	suspect Tamerlan Tsarnaev, killed in a shootout after the blast, has been buried at an location, policies Worcester Mass. Said
LexRank+SC+both	**Boston bombing** suspect Tamerlan Tsarnaev, killed in a shootout after the blast, has been buried at an location police said.
Ground Truth	**Boston bombing** suspect Tamerlan Tsarnaev has been buried at an undisclosed location

Impact of alpha and beta



(a) Impact of α



(b) Impact of β

Summary and Future Work

- We use a tweets related pipeline approach for compressive news highlights generation.
- We extend a dependency-tree based model to incorporate tweet information for sentence compression.
- We will explore more effective ways to incorporate tweets for sentence compression.
- We will study joint models to combine both sentence extraction and compression processes.
- We will investigate topic timeline generation via parallel dataset.

Reference

- Katja Filippova and Michael Strube. Dependency tree based sentence compression. NLGC’08
- Yang et. al. Social Context Summarization. SigIR’11
- Chen Li et. al. Improving Multi-documents Summarization by Sentence Compression based on Expanded Constituent Parse Tree. EMNLP’13
- Wei and Gao. Utilizing Microblogs for Automatic News Highlights Extraction. COLING’14
- Wei and Gao. Gibberish, Assistant, or Master? Using Tweets Linking to News for Extractive Single-Document Summarization. SigIR’15
- Wei et. al. Using Tweets to Help Sentence Compression for News Highlights Generation. ACL’15

Q&A

Using Tweets to Help Sentence Compression for News Highlights Generation

Zhongyu Wei[1], Yang Liu[1], Chen Li[1], Wei Gao[2]

[1]Computer Science Department, The University of Texas at Dallas, Texas, USA
[2]Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar

July 27th 2015 Beijing, China
The 53rd Annual Meetings of the Association for Computational Linguistics