Page Rank Applied to Personal Narratives' Most Reportable Event

Trudy Painter

Most Reportable Event

"The more reportable the most reportable event of a narrative, the greater justification for the automatic reassignment of speaker role to the narrator." (William Labov, University of Pennsylvania)

Labov, W. (1997). Some further steps in narrative analysis. *Journal of Narrative & Life History, 7*(1-4), 395–415.

What is most reportable event - explain

In 1997 defined as above by william labov

- Most important or interesting part of the story
- - like a turning point -
- how to keep the audience interested
- CITATION BELOW

Labov, W. (1997). Some further steps in narrative analysis. *Journal of Narrative & Life History, 7*(1-4), 395–415.

“Modeling Reportable Events as Turning Points in Narrative”

Jessica Ouyang and Kathleen McKeown: “the compelling event that is the nucleus of the story”

Ouyang, J., & McKeown, K. (2015, September). Modeling reportable events as turning points in narrative. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (pp. 2149-2158).

Researchers at Columbia University collected thousands of stories from reddit and labelled hundreds of them to create a corpus of short personal stories

They used a machine learning model to predict where the sentence with the most reportable event would be

Data

- Data sourced from story threads on reddit
  - What's your creepiest (REAL LIFE) story?
  - Your best "Accidentally Racist" story?
  - What are your stories of petty revenge?
- 4,178 collected stories -> 460 annotated stories
  - Interannotator agreement - 0.792

Ouyang and mckeown sourced personal stories from reddit

- Prompts above

Collected 4178 "stories"

- Some were inappropriate, or not even a story, or too short, or not a fully formed story

- 460 annotated stories
  - Used two annotators
  - Second annotator was just to check inter annotator agreement

Will use

This isn’t exactly creepy, but it’s one of the scariest things that’s ever happened to me. I was driving down the motorway with my boyfriend in the passenger seat, and my dad in the seat behind my own. My dad is an epileptic and his fits are extremely sporadic. Sometimes he goes extremely stiff and other times he will try to get out of places or grab and punch people. **Mid-conversation I felt his hands wrap around my throat as I was driving, pulling my head back and making it increasingly difficult to drive.** My boyfriend managed to help steer the car into the hard shoulder but it was one of the scariest experiences in my life.

Example of most reportable event from a story they used

Type	Metric Names
Syntactic	sentlength, vplength, lengthratio, sentdepth, vpdepth, depthratio wordlength, structcomplexity, wordformality, wordcomplexity
Semantic	cossimilarity, lssimilarity
Affectual	pleasantness, activation, imagery, subjectivity

Table 3: The fifteen metrics for change.

Trial	Precision	Recall	F-Score
Last sent. baseline	0.208	0.112	0.146
Heuristic baseline	0.107	0.333	0.162
No change*	0.146	0.378	0.211
Random baseline	0.185	0.586	0.281
Change only*	0.351	0.685	0.466
All features*	**0.398**	**0.745**	**0.519**

Their approach to modeling these sentences was based on syntactic and semantic changes in the sentences.

- Change in sentence length, word length, complexity
- Change in tone, pleasantness, activeness of of sentence

Paper's Approach

- Find **differences** in syntax and tone of sentences

My Approach

- Find **similarities** in vocabulary and key words of sentences

- Paper focused on the DIFFERENCES in the story's sentences that would predict a turning point
- I focused on what SIMILARITIES would predict the turning point of a story (vocabulary)
    - Aka a reference to important keywords that i hypothesized would occur in the introductory part of the story and as well as in the turning point

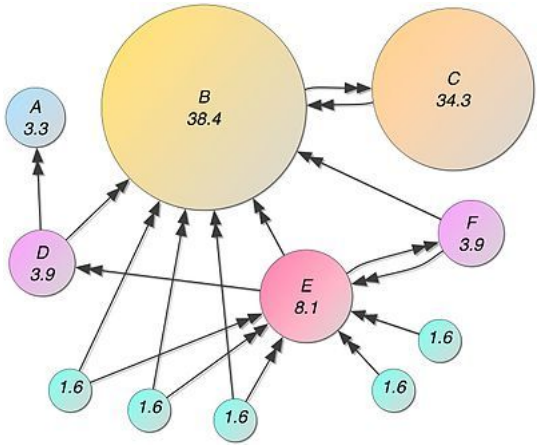
Paper also used machine learning (distant supervision)
I used pagerank which is a set predefined algorithm based off network graphs of the word embeddings

We both use similar word processing strategies to quantify the data

My Project

Apply Google's PageRank algorithm to vocabulary and keywords in the story.

I predicted that sentences with the most words in reference to the main idea of the story would contain the most reportable event.



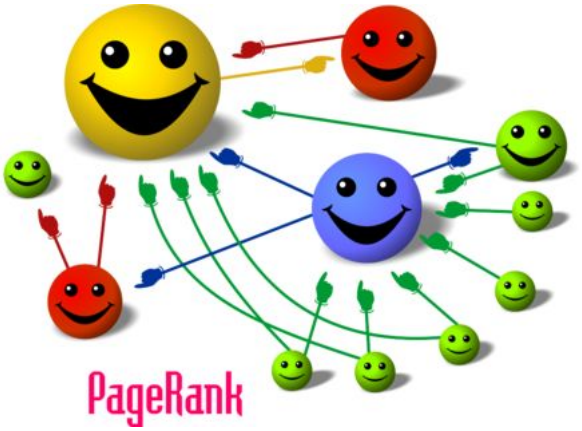
For my UROP project, I aimed to replicate their project with my own method for selecting the most reportable event

I had to figure out what was missing in their model and how to improve upon it.

Their model looked for changes - I predicted that a turning point in the story would contain keywords

How does PageRank work?

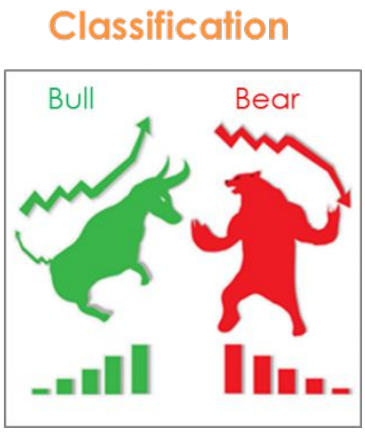
- Used by Google to sort the most relevant pages to the top of search results
- In explanation on the right, the size of each face is proportional to the total size of the other faces which are pointing to it
  - The biggest faces are most important or relevant

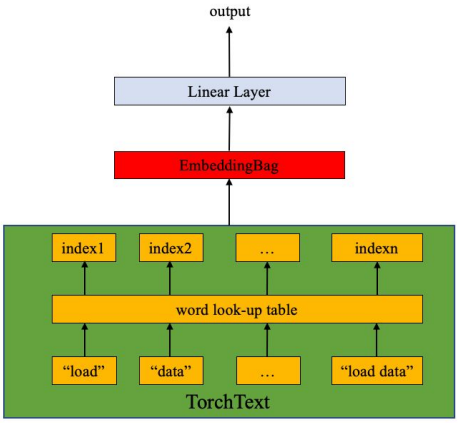


Words and similar words (due to word embeddings) that are used most commonly will be considered important words and appear in the middle of a similarity network

Preparation

- Teach myself about machine learning

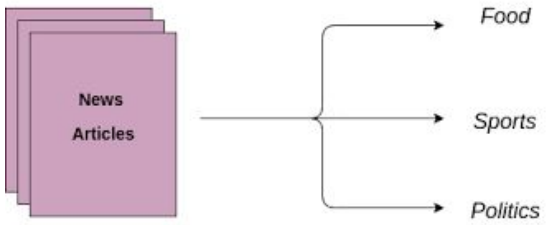




- Had never been exposed to machine learning before - my first time working with it
    - Belen provided me tutorials and research papers to learn more about the subject
    - Created basic models for image and story classification from news outlets

Inspiration

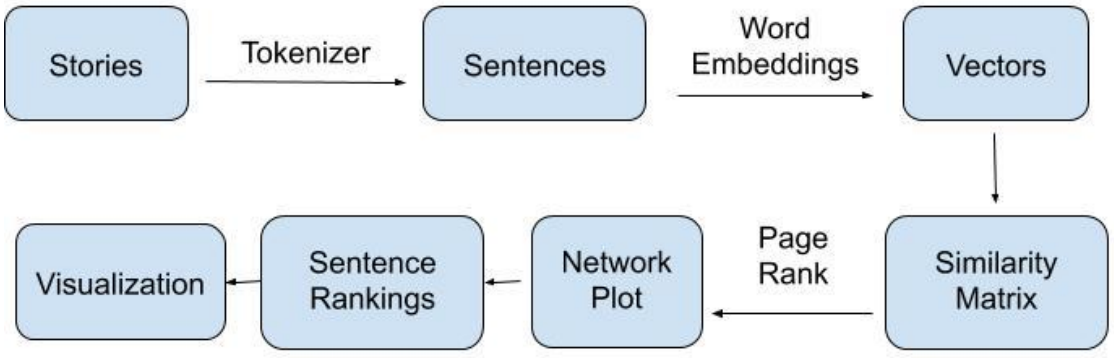
- Text classification tutorial made me realize how you can use word embeddings to determine most important topics
- I realized PageRank could be applied to most important keywords on a sentence-by-sentence level



- After doing text classification for news stories i realized that the key word classification could be used in a pagerank application

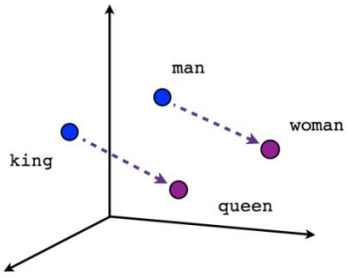
- Sentences could be ranked based off the amount of “important” content they contained (keywords)

Create Outline for Method

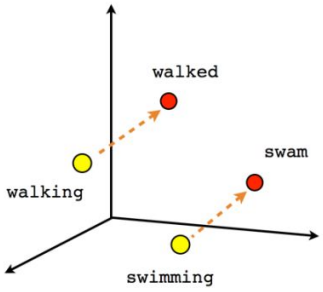


1. Stories
    a. Was originally going to try to create my own corpus of stories from audio
    b. Found youtube videos of strangers telling short stories in very similar length and genre to the columbia paper
    c. Could download all of the transcripts but formatting for that large of a corpus was nearly impossible. None of the transcripts had punctuation, and we used a tokenizer that was effective less than fifty percent of the time.
    d. Instead, used the annotated stories from the columbia experiment - thousands of stories with about 450 annotated.
2. Tokenizer to sentences- used same tokenizer to split up the stories into sentences (every story had periods)
3.

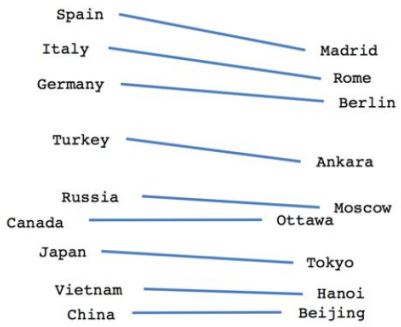
Word Embeddings



Male-Female



Verb tense

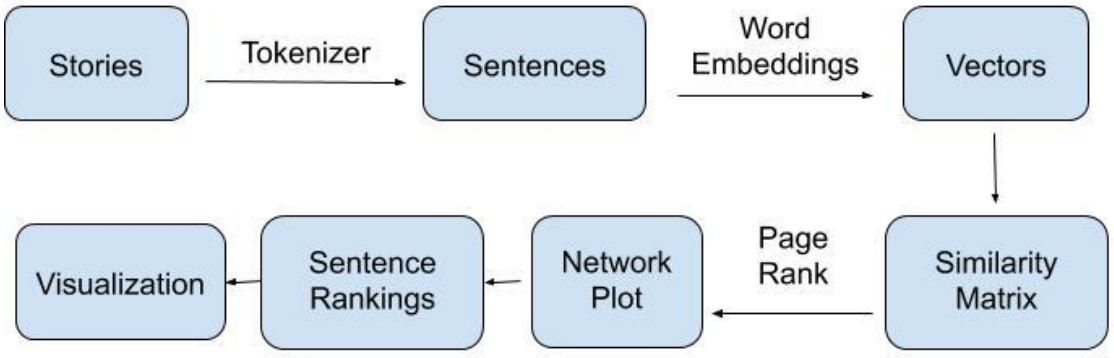


Country-Capital

Word embeddings are used to make sentences and words more qualitative to compare them

King - man = queen

Create Outline for Method



4. Similarities between the sentence vectors are then calculated and stored in a matrix
A dictionary of stopwords is used to remove common words like “the” and “and” from the sentences so that only important key vocabulary words are considered

Page rank is used to create a plot of the sentences - the sentences are the vertices and the edges are high similarity scores

Sentences towards the middle of the graph have more important words and are ranked as such.

Ranked three sentences because the columbia paper had an average of 2.5 ranked sentences

Visualization: created line graphs of sentence position based on ground truth and predicted - also network graph of the stories

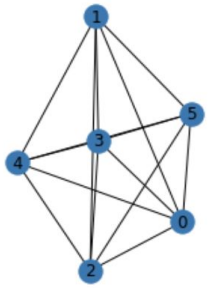
Example Results:

```
In [139]: printInfo(28)
```

GT | PREDICTED

```
28 | vfxbo.45.story| [3, 4]| [3, 5, 2] | ◯◯◓◉◒◓
```

```
In [134]: visualizeGraph(28)
```



0 When I was 5, I was at a grocery store with my mum and there was a homeless man sitting outside.
1 We went into the store and bought a loaf of bread to give him.
2 When we paid and left, my mum gave it to me to give to him and when I did, he tried to spit at me.
3 He started screaming that bread won't pay for beer and weed and yelled at me to go away and bring money next time.
4 My mum and I quickly left after that, but as a 5 year old, all I wanted to do was be kind, like I'd been taught.
5 It really upset 5-year-old me.

VISUALIZATION

28 is index of story in annotation

Vfxbo - prompt
45 - response number

3, 4 -- ground truth
3,5, 2 -- predicted

Line graph bottom half filled is ground truth -- top half is predicted

Bottom shows similarity graph

- Moderately succesful

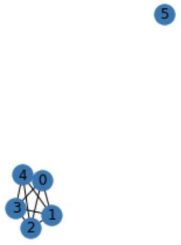
Bad Results

```
In [15]: printInfo(182)
```

GT | PREDICTED

```
182 | vfxbo.42.story|  [5] |  [2, 3, 4]  | ○○◓◓◓◒
Out[15]: ['182', 'vfxbo.42.story', [5], [2, 3, 4], '○○◓◓◓◒']

In [14]: visualizeGraph(182)
```



0 - Lent a boss $250 to buy this pos car.

1 - He had been diagnosed with ball cancer, I felt bad.

2 - And I had some extra scratch I could lend.

3 - He buys the car, which was a decent car and he got a great deal because it was from a relative.

4 - Sells it for $2500, stopped offering be jobs (it was independent contract) haven't heard from him since.

5 - Fuck.

5 is predicted story…

However… it is the only one not even in the graph

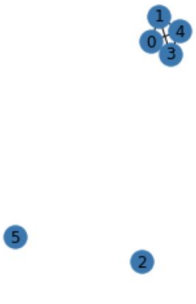
Quite frankly, i would argue it is the LEAST important sentence in the sentence - this may make you wonder if the data was coded incorrectly (like maybe inconsistent indexing)

Inconsistent Indexing

```
In [22]: printInfo(180)
```

GT | PREDICTED

```
180 | 1h7qys.44.story| [0] | [3, 4, 5] | ◒◯◯◓◓◓
Out[22]: ['180', '1h7qys.44.story', [0], [3, 4, 5], '◒◯◯◓◓◓']
In [21]: visualizeGraph(180)
```



0 - One time I took a dump that was so long, that even when trying to aim it down the hole at the bottom of the toilet it still stood up out of the water.

1 - When I pinched it off it had to fall one way.

2 - It fell forward and scraped across the backside of my ballsack as it fell.

3 - Tickled a bit, but was mostly startling more than anything.

4 - You don't expect to feel something on the back of your ballsack when you're taking a dump.

5 - Thought it could be a spider or something...but then I realized...

- Acknowledge the innapropriate-ness of this story. However, these are the types of stories that are represented in this dataset. Accurately represents reddit and genre of stories in it.
- Shows inconsistency in annotations
- Most importantly - the index zero is set as most important
    - So all indexes must start at zero then????

More Suspicious Results...

```
In [23]: printInfo(14)
         printInfo(22)
         printInfo(24)
```

```
14 | vfxbo.87.story|  [9] |  [2, 6, 7]  | ○○◓○○○◓◓○
22 | hzqqn.28.story|  [11] |  [9, 3, 5]  | ○○○◓○◓○○○◓○
24 | vfxbo.47.story|  [9] |  [8, 2, 7]  | ○○◓○○○○◓◓
```

Some indexes are out of range of the stories' length

So…. how can the annotations be trusted?

- This suggests

Inconsistent Indexing!

- Results show some stories were indexed starting at 0 and others at 1
- Can not fully trust annotations
  - Followed assumption they were indexed at 0 for experimental purposes

Overall Results

- Best run was from random sample of 25
  - Removing first two predicted sentences

a. Precision - 27%
   i. (accurate recovered sentences)/(total recovered sentences)
b. Recall - 18.8%
   i. (accurate recovered sentences)/(total MRE sentences)

INDEX	STORY	GT	PREDICTED
2447	hzqqn.66.story	3	4 3 2
4765	12he9h.2.story	13	15 6 12
1497	pwlhs.26.story	6 7	3 9 8
4178	oeo0h.77.story	4 5	5 3 6
631	11urgx.4.story	2 3 4	3 5 4
2520	uie6f.16.story	5	17 10 4
1263	vfxbo.79.story	13	3 8 16
1064	vfxbo.22.story	10	12 5 6
2165	1h7qys.27.story	17 18 8 9	7 14 11
4634	27fdf1.1.story	6 8	3 5 2
848	27fdf1.15.story	11 13 4	8 2 10
3344	1air31.37.story	13 14 15 16	13 17 12
4088	1air31.29.story	16 18 7	12 8 4
1856	1h7qys.44.story	0	4 3 5
1871	uie6f.25.story	8	5 4 10
356	1air31.21.story	2 3	5 4 3
3257	oeo0h.35.story	10 11 12 7 8 9	2 10 9
2508	vfxbo.34.story	4 5	5 6 4
1239	1r034d.24.story	10 11	9 7 11
271	hzqqn.28.story	11	3 5 9
3314	y9dex.40.story	4 5	2 17 14
1540	vfxbo.120.story	18	19 10 13
2228	vfxbo.59.story	3 4	3 5 4

RESULTS

Algorithm vs Random Results

```
In [64]: get_pres_and_recall(selected_results40)
         pres:0.20015396458814472 recall:0.3051643192488263

In [65]: get_pres_and_recall(selected_results20)
         pres:0.22128851540616246 recall:0.3640552995391705

In [66]: get_pres_and_recall(results_df)
         pres:0.19420289855072465 recall:0.27714581178903824

In [67]: get_pres_and_recall(random_results_df)
         pres:0.17768595041322313 recall:0.24457083764219234
```

- First two results: stories shorter than 40 and 20 sent.
- Third results: overall results for entire dataset using PageRank
- Fourth results: random index selection

Algorithm is effective on shorter stories

- Slightly better than random

Interpretations

- Model is better than random
- Often, algorithm picked longer sentences with lots of “important” words
- Skeptical of data

Mre was defined as turning point - not necessarily summary
Sometimes mre could be super short - story about ufo and aliens - labeled mre about ufo “it moved” not picked up by page rank

Data of the columbia paper

- Skeptical of how it was annotated -- inter annotator agreement didn’t seem to line up with the numbers in their paper

Further study

1. Could this approach be added to the syntatic/semantic change approach?
    a. Would a combination of both lead to better results?
2. Where are other places to pull short personal narratives?
3. Creating better data visualizations could lead to better understanding of data.
4. Are there better ways to standardize the dataset?

Sometimes didn't know exactly where/how the data was being processed.