Summarization of Twitter Accounts

Final Report

Michael Alvin
College of Engineering
University of Michigan
Ann Arbor, MI
mchlalvn@umich.edu

Priyanka Shanmugasundaram
School of Information
University of Michigan
Ann Arbor, MI
pshanmu@umich.edu

John Wyman
School of Information
University of Michigan
Ann Arbor, MI
jwyman@umich.edu

ABSTRACT

Since its creation in 2006, Twitter has been consistently one of the most popular social media platforms. As a result, the platform is filled with an abundance of information. The enormous size of Twitter data calls a need for summarization systems. This task is called multi-tweet summarization. Multi-tweet summarization is applicable to important issues, and recent research has ranged from the summarization of trending topics to summarization of live disastrous events. In this research project, we explore the task of summarizing Twitter accounts. We built three separate summarization models: a graph-based model, a cluster-based model, and a random model. Our results show that graph-based model is most effective in this task.

KEYWORDS

Information Retrieval, Text Summarization, Extractive Summarization, LexRank, Latent Dirichlet Allocation, Twitter

1 INTRODUCTION

1.1 Problem

In recent years, the rise of technology and the Internet have brought unprecedented changes. One of these changes is the introduction of social media platforms where people can create and share content or to participate in social networking. Some of the more popular social media platforms include Facebook, Reddit, Snapchat, Twitter, etc. Most people use at least one of these platforms, and in today's world it would be uncanny for a person to not use social media. Many important figures even use social media as their main mode of communication, including singer Taylor Swift, athlete LeBron James, and even the President of the United States Donald Trump.

In short, these platforms allow people to connect and enable people to follow certain important figures. However, there are so many posts, threads, snaps, tweets in today's digital world. For instance, there are over 330 million active Twitter users and 500 million tweets are sent each day. The sheer number of accounts on Twitter makes finding an account to follow that is relevant to a user's interests all the more difficult.

For this reason, our project revolves around building a Twitter account summarization system. In this project, we will create Twitter account summarizer systems that extracts the most important tweets in a profile and selects the most important topics related to the profile. In this project, we employ a cluster-based approach and a graph-based approach to this task and compare the results of the different approaches. This application is useful for those who want to recommend a Twitter account or are interested in a particular Twitter handle and need a simple way to explain what the account focuses on. This project can be used to summarize other text-based social media platforms and be used as a great pre-processing step in analyzing social media platforms.

2 PREVIOUS WORK

2.1 Multi-tweet Summarization

There are several research papers exploring the topic of multi-tweet summarization. [1] explores the summarization of topics (such as iPads, holidays, Windows, etc.) from tweets of different users. The research employs a graph-based approach, augmented with social features such as follower numbers, retweeted times and other features such as readability. [2] explores the summarization of events through a classification (situational, non-situational) then summarization approach, using Integer Programming to maximize the coverage of content words in the tweets included in the summary.

2.2 Twitter

There are also substantial research papers revolving Twitter data in general. [3] explores the topological characteristics of Twitter and its power as a new medium for information sharing. There are also papers exploring how Twitter is used by the members of the United States Congress [4] and police departments in large United States cities [5]. In short, Twitter is an influential platform and contains enormous, valuable data.

3 METHODS

3.1 Data

The data used to conduct the evaluation was Twitter data collected from the Twitter API. It was in typical Twitter data JSON formats, and collected via a development Twitter account created specifically for this project. Python's Tweepy library was used to collect the data from Twitter API.

For this project, we identified a group of diverse, prolific tweeters to evaluate our models: Justin Bieber, Katy Perry, Gordon Ramsay, Donald Trump, Elon Musk, Wendy's, and Neil DeGrasse Tyson. From the 8 Twitter accounts, we collected their most recent 1,000 tweets. A sample Tweet data object is shown below.

We limited 1,000 tweets per user, because we had to manually hand-select the most relevant tweets for each user. We believe increasing the number of tweets per user would compromise the quality of our selections and would prevent us to evaluate in a timely manner.

{
"created_at": "Wed Nov 28 19:33:041044949188788682800 +0000 2018",
"id": 1067863966829801500,
"text": "We're back. Season 3 premiere. Monday February 4th 8:30 pm/7:30 central. @ManWithAPlan @CBS #ManWithAPlan #CBS https://t.co/XWtTrR5uzR",
…
}

3.2 Data Pre-processing

All "retweet" data was filtered to create a more accurate corpus for the handle. Retweets are when a Twitter user forwards or repeats something written by another user. For the sake of this project's goals, this data was less applicable than text actually entered by a user.

3.3 Algorithms

In this project, we implemented three models: base model, clustered-based model, and graph-based model. All three models consider the tweets from each user and output the top 50 (5% of 1,000) most representative tweets.

3.3.1 Base Model, Random. The base model simply selects 50 random tweets as the summary for each Twitter user. This trivial model serves as our base model and we predict the other two models would outperform this model.

3.3.2 Cluster-based Model, Latent Dirichlet Allocation. The cluster-based model used Latent Dirichlet Allocation [6] to generate a topic model from each user's set of tweets. In this project, the topic modelling algorithm generated five topics and the top documents from each topic were gathered.

Algorithm 1 Latent Dirichlet Allocation
Input: 1,000 twitter tweets, **Output:** 5 topics, 10 most relevant twitter tweets from each topic

---

1 For each word w in a tweet *t*, calculate its tf-idf vector representation.
2 Assume there are k topics across all documents.
3 Perform Latent Dirichlet Allocation algorithm on tf-idf vector.
4 For each topic, sort documents and output the most relevant tweets.

3.3.3 Graph-based Model, LexRank. The graph-based model, LexRank, builds a sentence network out of a user's set of tweets. This model [7] was inspired by the well-known PageRank algorithm and have proved to be effective in summarization tasks.

Algorithm 2 LexRank
Input: 1,000 twitter tweets, **Output:** 50 most representative twitter tweets

---

1 For a tweet *t*, calculate its tf-idf vector representation.
2 Build a graph *g*, where each vertex *v* represents a tweet and each edge *e* represents cosine similarity between any two tweets.
3 Filter edges *e* that are above a certain threshold and zero the edges below the threshold.
4 Perform PageRank algorithm on the graph *g*.
5 Sort the edges *e* by the PageRank score and output the highest scored tweets.

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}}$$

Formula 1: Cosine Similarity

$$PR(p_i) = \frac{1-d}{N} + d \sum_{p_j \in M(p_i)} \frac{PR(p_j)}{L(p_j)}$$

Formula 2: PageRank

4 EXPERIMENT

4.1 Evaluation

To evaluate our models, we manually selected 50 tweets from each of the eight Twitter account that we considered most representative of the account's activity. These hand picked 50 tweets served as the gold summary to be compared with the 50 tweets our models output.

Per our research, ROUGE is the main evaluation metric for summarization systems. There are many variations of ROUGE: ROUGE-n, ROUGE-L, etc. In this project, given the short, orderless nature of tweets, we chose to evaluate using ROUGE-1 and ROUGE-2, which calculates the percent of unigrams (ROUGE-1) and bigrams (ROUGE-2) from the gold summary show up in the system summary. We used an open-sourced Google library[1] that implements the ROUGE evaluation and calculates our ROUGE-1, ROUGE-2 scores.

$$\text{ROUGE-}n = \frac{\sum_{C \in RSS} \sum_{gram_n \in C} \text{Count}_{match}(gram_n)}{\sum_{C \in RSS} \sum_{gram_n \in C} \text{Count}(gram_n)}$$

Formula 3: ROUGE-n

4.2 Results

The ROUGE-1, ROUGE-2 scores of the three systems are shown below. The best performing summarization system for each Twitter account is highlighted.

	LexRank	**LDA**	**Random**
elonmusk	0.4929	**0.5753**	0.4285
barackobama	**0.6337**	0.6195	0.6082
realdonaldtrump	**0.6449**	0.6169	0.5688
justinbieber	**0.5348**	0.4693	0.5234
neiltyson	**0.4952**	0.4646	0.4916
wendys	0.6369	0.5479	**0.6849**
gordonramsay	**0.6084**	0.5731	0.5557
katyperry	**0.6060**	0.5558	0.4839

Figure 1: ROUGE-1 Scores

	LexRank	**LDA**	**Random**
elonmusk	0.2053	**0.2713**	0.1204
barackobama	**0.3087**	0.2914	0.2209
realdonaldtrump	**0.2824**	0.2707	0.1935
justinbieber	**0.2616**	0.1963	0.1861
neiltyson	0.1050	0.0971	**0.1179**
wendys	0.4515	0.3302	**0.5331**
gordonramsay	**0.2960**	0.2702	0.2549
katyperry	**0.2311**	0.2023	0.1227

Figure 2: ROUGE-2 Scores

[1] https://github.com/google-research/google-research/tree/master/rouge

---

Furthermore, below we can compare the quality of the summary results from the LexRank model and the random model. The LexRank tweets conveys more information and paints a picture of what Elon Musk's Twitter account is about, whereas the random tweets do not tell us much.

We're creating this alloy at Tesla. Not a problem to create a lot of it, but we'll need to come up with new body manufacturing methods, as it can't be made using standard methods.
We'll aim to have it come out same time as truck. Two seater electric ATV designed to work with Cybertruck will be fun! Electric dirt bikes would be cool too. We won't do road bikes, as too dangerous. I was hit by a truck & almost died on one when I was 17.
I'd be way too embarrassed to put that on a Tesla. It's like a kid's drawing.
SpaceX engine production is gearing up to build about a Raptor a day by next year, so up to 365 engines per year. Most will be the (as high as) 300 ton thrust (but no throttle & no gimbal) variant for Super Heavy. Cumulative thrust/year could thus be as high as 100,000 tons/year.
No, in the beginning, assuming you even make it there alive, Mars will be far more dangerous & difficult than Earth & take decades of hard labor to make self-sufficient. That's the sales pitch. Want to go?

Figure 3: LexRank Top-5 Tweets Summary

We can prob reduce width by an inch & maybe reduce length by 6+ inches without losing on utility or esthetics. Min height is below 75 inches when air suspension set to low. Will post exact number soon.
🤣🤣
Wow, 2011 seems like eons ago! With fairing recovery, Falcon is ~80% reusable, but reflight takes several days & requires boats. Starship will be fully reusable with booster reflight possible every few hours & ship reflight every 8 hours. No boats needed.
Without a fully & rapidly reusable orbital rocket, humanity will never be a multiplanet species
187k

Figure 4: Random Top-5 Tweets Summary

5 DISCUSSION

5.1 Discussion

In general, the graph-based model is most effective in summarizing Twitter accounts. The results show that LexRank outscores Random in 13 of 16 tests, LDA outscores Random in 11 of 16 tests, and LexRank outscores LDA in 14 of 16 tests. Therefore, the results demonstrate that Twitter accounts can be summarized in an effective manner using LexRank.

The results of this project suggests that the lower the score for a handle's evaluation, the more diverse the group of subjects are that the person tweets about. Likewise, the higher the score, the less diverse the topics. For example, the Wendy's account was an outlier and had high ROUGE scores. However, we also noticed that this account had a lot of repetition.

5.2 Future Work

For future work, there are several directions we can pursue:

- Reflecting on what we learned in the process, we identified some additional opportunities to evaluate Twitter traffic. Particularly in the task of manually labeling the data used for the Rouge analysis. We could improve the quality of the scoring by increasing the number of people who manually score the tweets. The time allotted to the project didn't allow for a large sample size, but it would improve the quality of the research through increased quality in the manual labeling. Lastly, expanding beyond 8 twitter handles would increase the breadth of samples, and the veracity of our results.
- In addition, we may improve our graph-based LexRank model in several ways. We observed that sometimes the summarizer would output multiple tweets regarding the same topic. Instead of simply selecting the highest page ranked tweets, we should also consider the similarity between selected tweets so that no two selected tweets are too similar. Furthermore, we may augment our LexRank model by adding more features, including social features such as number of retweets and/or tweet features such as whether the tweet contains a video. Instead of simply using the page rank scores, we can take a linear combination of the page rank scores and any additional features we think may be important.

5.3 Code and Demo

Our source code is publicly accessible at https://github.com/jwymanumich/SI650Team.
Our demo is publicly accessible at https://youtu.be/Nmn4nV7pRnGU.

ACKNOWLEDGMENTS

We are grateful to Professor Qiaozhu Mei, Professor Wei Ai, Shiyan Yan, and Xinyan Zhao for their guidance and inspiration.

REFERENCES

[1] Xiaohua Liu, Yitong Li, Furu Wei, and Ming Zhou. 2012. Graph-based multi-tweet summarization using social signals. In Proceedings of COLING 2012.

[2] Koustav Rudra, Subham Ghosh, Niloy Ganguly, Pawan Goyal, and Saptarshi Ghosh. 2015. Extracting situational information from microblogs during disaster events: a classification-summarization approach. In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management 2015.

[3] Kwak, Haewoon, Changhyun Lee, Hosung Park, and Sue Moon. 2010. What is Twitter, a social network or a news media? In Proceedings of the 19th international conference on World wide web 2010.

[4] Jennifer Golbeck, Justin M. Grimes, and Anthony Rogers. 2010. Twitter use by the US Congress. Journal of the American Society for Information Science and Technology 61, no. 8 (2010): 1612-1621.

[5] Thomas Heverin and Lisl Zach. 2010. Twitter for city police department information sharing. Proceedings of the American Society for Information Science and Technology 47, no. 1 (2010): 1-7.

[6] Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." Journal of Machine Learning Research 3, no. (Jan 2003): 993-1022.

[7] Günes Erkan and Dragomir R. Radev. 2004. Lexrank: Graph-based lexical centrality as salience in text summarization. Journal of artificial intelligence research 22 (2004): 457-479.