

facebook

Likes Estimator for Major News Publishers' Pages

2015 Fall Data Science Final Project
Advisor: Ming-Syan Chen
Hank Yao, Yi-Te Hsu, Janet Sung

Outline

- Motivation
- Resources
- Feature Extraction
- Data Cleaning
- Visualization
- Model Construction
- Future Work

· Motivation · Resources · Feature Extraction · Data Cleaning · Visualization · Model Construction · Future Work

Motivation

how to get more likes on a facebook page post

全部 影片 圖片 新聞 地圖 更多 搜尋工具

約有 1,260,000,000 項結果 (搜尋時間: 0.58 秒)

Here's a Quick Way to Get More Likes on Your Facebook Page
<https://www.postplanner.com/get-more-likes-fans-facebook...>
 The #1 question I bet you want answered: **How do you get more likes on your Facebook Page?** My answer: use the Viral Photos feature in Post Planner to post ...
 Give Me 5 Minutes a Day and I ... - See Post Insights

Here's a *FREE* New Way to Get More Likes on Facebook ...
<https://www.postplanner.com/free-new-way-get-more-likes...>
 2014年1月4日 - Here's how you can get FREE Likes on your posts using a new tactic ...
 how to get more Likes on your Facebook posts -- not just your page.

How to Get More Likes on Your Facebook Posts [Infographic]
<https://www.postplanner.com/how-to-get-more-likes-on-fa...>
 How to Get More Likes on Your Facebook Posts [Infographic]. Facebook Marketing ...
 I'm not a fan of saying there is a "magic" time of day to post to your page.

10 Ways to Get More Likes for Your Brand's Facebook Page ...
[www.meltwater.com > Blog](http://www.meltwater.com/Blog)
 2015年7月10日 - This blog post discusses how to get more likes for your Facebook Business Page.

Is there a scientific method to make the job of a page content creator easier?

Ex.

"Something special appears to happen in the brain while the body floats."

Submit

Recommended Post Time: 22:00 pm

Recommended Content:

Better to include a location or name.

Hashtag needed.

► Motivation · Resources · Feature Extraction · Data Cleaning · Visualization · Model Construction · Future Work

Resources

- Scripting Language
Python, Java
- API
Facebook Graph API, Stanford CoreNLP
- Data Sources
Facebook pages of CNN, BBC, TIME, Huffington Post, The Economist, New York Times, Wall Street Journal

· Motivation ► Resources · Feature Extraction · Data Cleaning · Visualization · Model Construction · Future Work

Resources: Facebook Graph API

Graph API is a primary way to read and write Facebook social graph. It requires an access token and page ID.

1. Facebook Graph API Explorer

<https://developers.facebook.com/tools/explorer/>

2. Lookup-ID.com

<https://lookup-id.com/>

Brute Force: Python BeautifulSoup

· Motivation ► Resources · Feature Extraction · Data Cleaning · Visualization · Model Construction · Future Work

Resources: Facebook Graph API

```
import facebook

import datetime
from time import strftime
token = '1085103088207726|CAACEdEose0cBAAK1TdGFFQTPaVEvPTBVKrL4RxTkf0CLottIRSfxCMJtur4bZBprIrAwnwPdI7kY5en24301jLEr27bDcl
graph=facebook.(token)
#ID = '5281959998' #The New York Times 10,219,642
#ID = '5550296508' #CNN 19,571,073
#ID = '228735667216' #BBC News 25,498,527
#ID = '18468761129' #The Huffington Post 5,478,619
#ID = '10606591490' #TIME 8,017,895
#ID = '8304333127' #The Wall Street Journal 4,295,516
#ID = '6013004059' #The Economist 5,095,942
page={'NYT': '5281959998', 'CNN': '5550296508', 'BBC': '228735667216', 'Huffington': '18468761129', 'TIME': '10606591490', 'WSJ':

#If True:
for keys in page:
    summ =0;
    ID = page[keys]
    #page_information
    ...
    page_info = graph.get_object(ID)
    page_likes = page_info['likes']
    ...
    page_query = "SELECT page_id, name, page_url, fan_count FROM page WHERE " + "page_id = " + ID#nytimes "# + page_url
    page_fql = after_fql = graph.fql(page_query)
    page_likes = page_fql['data'][0]['fan_count']
```

NOTICE!

Query function and object names
vary for Graph 1.0 and Graph 2.0

· Motivation ► Resources · Feature Extraction · Data Cleaning · Visualization · Model Construction · Future Work

Resources: Stanford CoreNLP

CoreNLP

version 3.6

- Introduction
- Usage
- Annotators**
- Summary
- Annotator dependencies
- Tokenization
- Sentence Splitting
- Lemmatization
- Named Entity Recognition
- RegexNER (Named Entity Recognition)
- Constituency Parsing

Stanford CoreNLP – a suite of core NLP tools

Table of Contents

- About
- Download
- Human languages supported
- Programming languages and operating systems
- License
- Citing Stanford CoreNLP in papers

About

Stanford CoreNLP provides a set of natural language analysis tools. It can give the base forms of words, their parts of speech, whether they are names of companies, people, etc., normalize dates, times, and numeric quantities, and mark up the structure of sentences in terms of phrases and word dependencies, indicate which noun phrases refer to the same entities, indicate sentiment, extract open-class relations between mentions, etc.

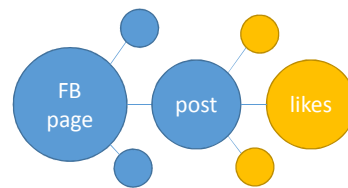
Java command line usage recommended. (There is some problem with its Python package.)

```
java -cp stanford-corenlp-VV.jar:stanford-corenlp-VV-models.jar:xom.jar:joda-time.jar:jollyday.jar:ejml-VV.jar -Xmx2g edu.stanford.nlp.pipeline.StanfordCoreNLP -file <YOUR INPUT FILE>
```

• Motivation ▶ Resources • Feature Extraction • Data Cleaning • Visualization • Model Construction • Future Work

Feature Extraction

The screenshot shows a Facebook post from the CNN page. The post text is "5-year-old Jaylie Vega's body was found after he went missing earlier this week". The post has 1,182 likes and 179 shares. The screenshot highlights the fan count, the post text, and the like count.



Following nodes are retrieved as features:

- fan_count: for normalization
- created_time
- message: NLP input
- likes.count

• Motivation • Resources ▶ Feature Extraction • Data Cleaning • Visualization • Model Construction • Future Work

Data Cleaning

Data: 27014 objects (train:test=9:1)

Attributes: 10 features

Content Related

- words – total number of words
- location – content with place
- person – content with name
- emoji – with emoji
- hashtag – with hashtag
- sent – sentiment level

Time Related

- time – post time (0-23)
- day – post day (0-6)
- thanksgiving – days after thg.
- paris_attack – days after pa.

Class: continuous or categorical

- rank – five category (1-5)
- n_likes – normalized number of likes

· Motivation · Resources · Feature Extraction ▶ Data Cleaning · Visualization · Model Construction · Future Work

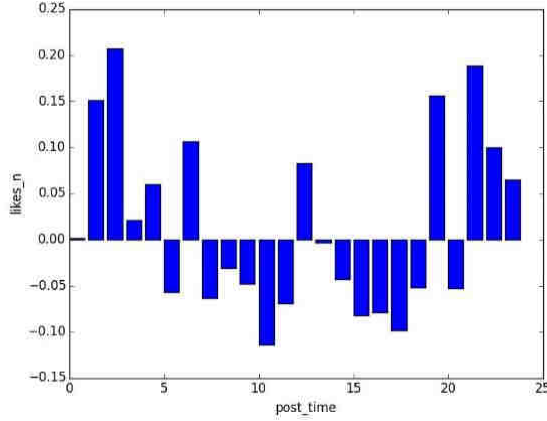
Data Cleaning: PCA

Feature	1 st Component	Feature	1 st Component
words	0.9976	location	0.0139
paris_attack	-0.0444	person	0.0109
thanksgiving	0.0443	day	-0.0103
sentiment	-0.0159	emoji	0.0077
time	-0.0146	hashtag	0.0016

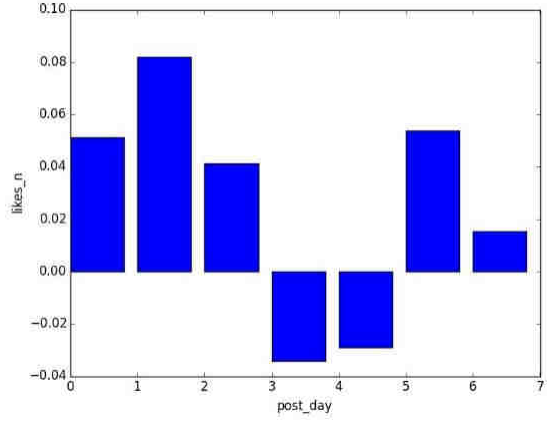
PCA suggests that time-related features have more impact on the number of likes.

· Motivation · Resources · Feature Extraction ▶ Data Cleaning · Visualization · Model Construction · Future Work

Visualization: posting time



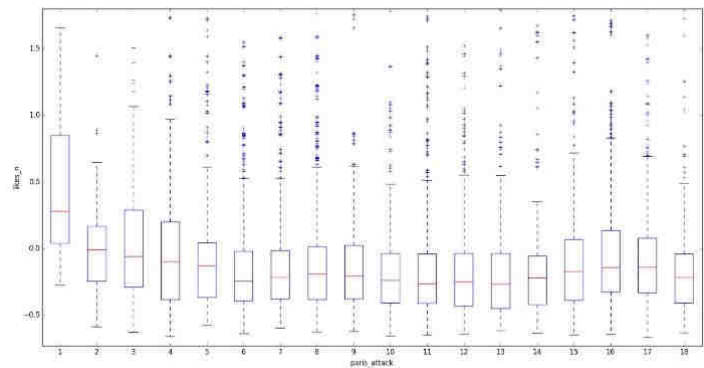
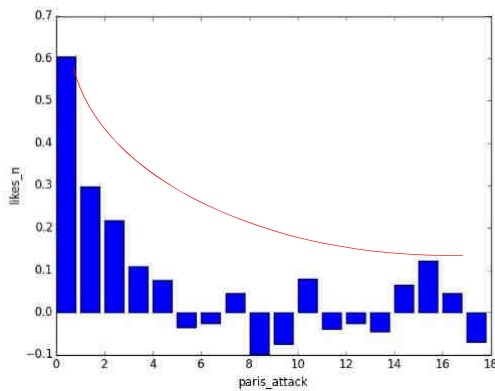
Posts created during working hours (7-12, 13-18) obtain less likes.



Posts created on Wednesday and Thursday obtain less like.

· Motivation · Resources · Feature Extraction · Data Cleaning ► Visualization · Model Construction · Future Work

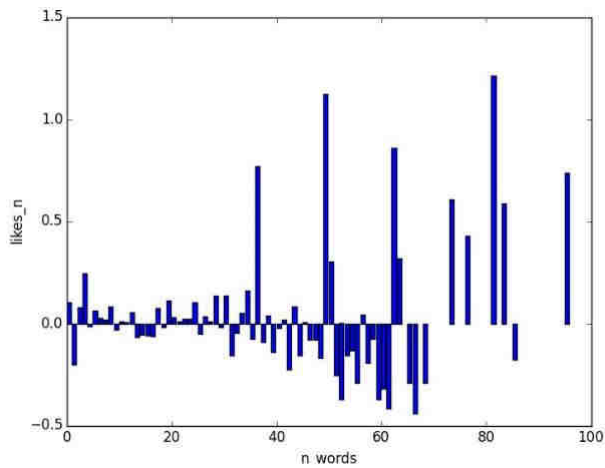
Visualization: event and likes



The closer of the created time to the Paris attack, the more likes. Possibly because people pay more attention to news during that period. However, variance of posts created within 1 day is huge because there were news of other topics which cancel out the likes for Paris attack related posts.

· Motivation · Resources · Feature Extraction · Data Cleaning ► Visualization · Model Construction · Future Work

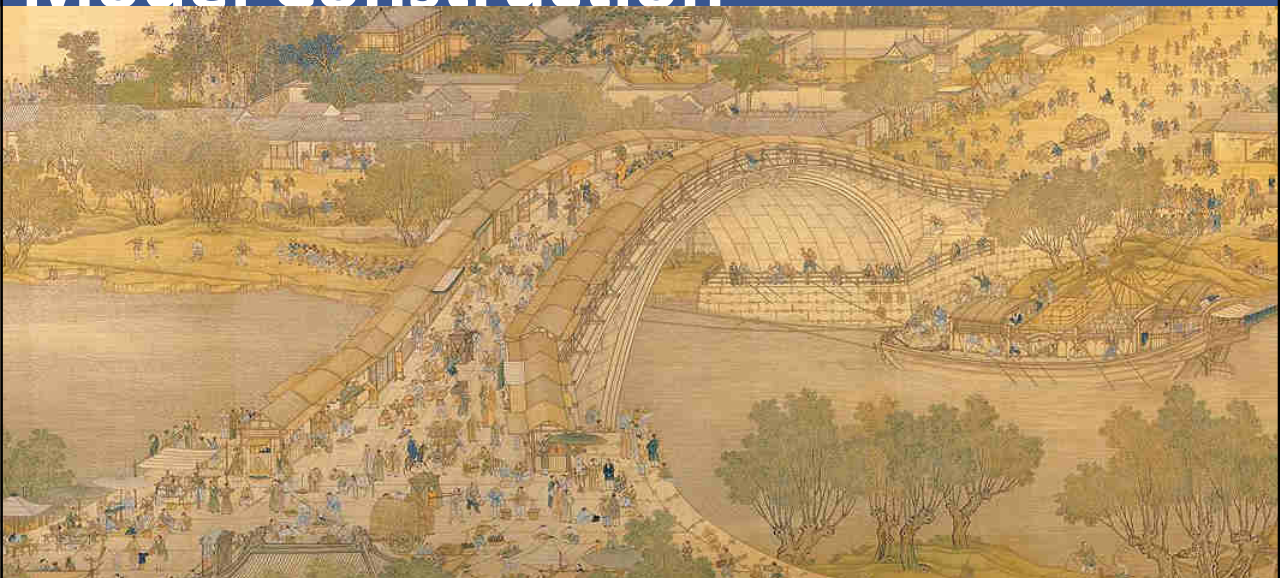
Visualization: number of words



Posts with either short content ($4 < \text{words} < 15$) or insightful information (> 75) are likely to get more likes.

· Motivation · Resources · Feature Extraction · Data Cleaning ▶ Visualization · Model Construction · Future Work

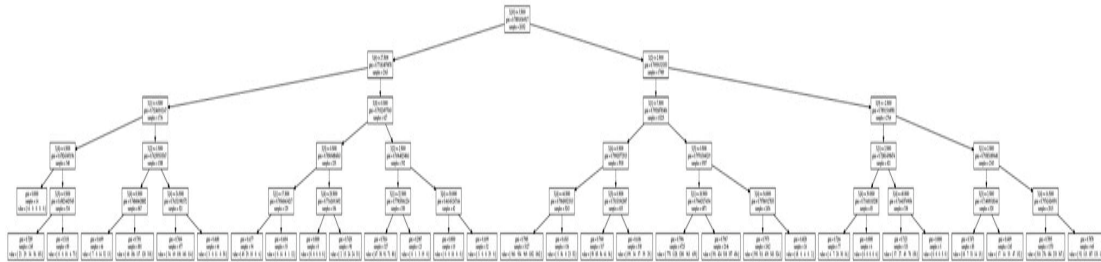
Model Construction



· Motivation · Resources · Feature Extraction · Data Cleaning · Visualization ▶ Model Construction · Future Work

Model Construction

Naïve Bayes: Gaussian (25.25%), Multinomial (23.30%), Combined (24.34%)
 Decision Tree: Decision Tree (92%), Random Forest (95%)



· Motivation · Resources · Feature Extraction · Data Cleaning · Visualization ▶ Model Construction · Future Work

Model: Decision Tree

Feature Importance

- words : 0.305673
- time : 0.258793
- thanksgiving : 0.074156
- day : 0.072380
- paris_attack : 0.070528
- emoji : 0.057877
- location : 0.052961
- sentiment : 0.047749
- person : 0.030926
- hashtag : 0.028957

- Using only top 5 features: accuracy drops to 85%.
- With or without hashtag does not have much impact on the number of likes.
- Result of sentiment is biased since there is no negative post.
- Best word range: [6,24]

To obtain the most likes,
 1-2 sentences are enough.
 Post on 20:00-22:00.
 Avoid Wednesday and Thursday.

· Motivation · Resources · Feature Extraction · Data Cleaning · Visualization ▶ Model Construction · Future Work

Future Work

- Better models
Transform the features to be suitable to some models
- More data
Data from a longer time-span
- Image and video content
- Extend to other applications
Personal post likes,
Chinese pages, etc.



· Motivation · Resources · Feature Extraction · Data Cleaning · Visualization · Model Construction ▶ Future Work

facebook Likes Estimator for Major News Publishers' Pages

Result Summary

- Accuracy: 95%
- Emoji, hashtag and name are not important
- Number of words and post time dominates
1-2 sentences, ~~Wed. and Thu.~~, evening and night

Thanks for your attention.

Questions and discussion are welcomed.