

Thesis Proposal

Yi Cui

ID Number: 15307130001

Fudan University

Oct 9, 2018

Outline

- 1 Data Selection
- 2 Data Exhibition
- 3 Data Analysis
- 4 Future Work

Outline

- 1 Data Selection
- 2 Data Exhibition
- 3 Data Analysis
- 4 Future Work

Motivation

- Economics combines with data science
 - (*Python, R, MATLAB* are my main tools to analysis data)

Motivation

- Economics combines with data science
 - (*Python, R, MATLAB* are my main tools to analysis data)
- First dataset is about Weibo, which provides social networking services. Also, its data are open-sourcing

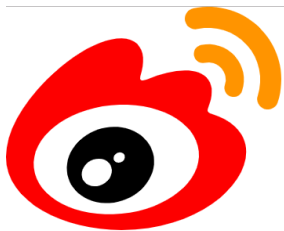
Motivation

- Economics combines with data science
 - (*Python, R, MATLAB* are my main tools to analysis data)
- First dataset is about Weibo, which provides social networking services. Also, its data are open-sourcing
- Second dataset is about finance: Lending Club Dataset

Motivation

- Economics combines with data science
 - (*Python, R, MATLAB* are my main tools to analysis data)
- First dataset is about Weibo, which provides social networking services. Also, its data are open-sourcing
- Second dataset is about finance: Lending Club Dataset
 - <https://www.kaggle.com/wendykan/lending-club-loan-data>

- "Weibo" is the Chinese word for "microblog":



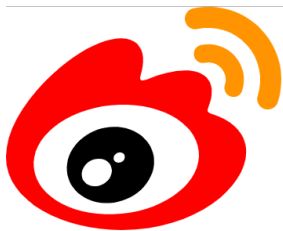
Weibo
Created by Zhidong Wang
August 14th, 2009

- "Weibo" is the Chinese word for "microblog":
 - In August 2009, Sina Weibo was launched



Weibo
Created by Zhidong Wang
August 14th, 2009

- "Weibo" is the Chinese word for "microblog":
 - In August 2009, Sina Weibo was launched
 - In early 2010, launched the API open platform



Weibo
Created by Zhidong Wang
August 14th, 2009



Weibo
Created by Zhidong Wang
August 14th, 2009

- "Weibo" is the Chinese word for "microblog":
 - In August 2009, Sina Weibo was launched
 - In early 2010, launched the API open platform
 - In April 2013, Alibaba acquired an 18% stake in Sina Weibo



Weibo
Created by Zhidong Wang
August 14th, 2009

- "Weibo" is the Chinese word for "microblog":
 - In August 2009, Sina Weibo was launched
 - In early 2010, launched the API open platform
 - In April 2013, Alibaba acquired an 18% stake in Sina Weibo
 - In April 2014, weibo was listed on Nasdaq



Weibo
Created by Zhidong Wang
August 14th, 2009

- "Weibo" is the Chinese word for "microblog":
 - In August 2009, Sina Weibo was launched
 - In early 2010, launched the API open platform
 - In April 2013, Alibaba acquired an 18% stake in Sina Weibo
 - In April 2014, weibo was listed on Nasdaq
 - In December 2016, Sina Weibo posted 313 million monthly active users and 139 million daily active users



Weibo
Created by Zhidong Wang
August 14th, 2009

- "Weibo" is the Chinese word for "microblog":
 - In August 2009, Sina Weibo was launched
 - In early 2010, launched the API open platform
 - In April 2013, Alibaba acquired an 18% stake in Sina Weibo
 - In April 2014, weibo was listed on Nasdaq
 - In December 2016, Sina Weibo posted 313 million monthly active users and 139 million daily active users
 - As of Q2 2018, Sina Weibo has over 431 million monthly active users, with surging stocks, advertising sales, revenue and total earnings in the quarter.

Lending Club Data

- Lending Club is a platform intermediary service that provides P2P loans to the market:



Lending Club
IPO in 2014
Created in 2006

Lending Club Data

- Lending Club is a platform intermediary service that provides P2P loans to the market:



- In December 2014, Lending Club was listed on the NYSE, becoming the largest technology stock IPO of the year.

Lending Club
IPO in 2014
Created in 2006

Lending Club Data



Lending Club
IPO in 2014
Created in 2006

- Lending Club is a platform intermediary service that provides P2P loans to the market:
 - In December 2014, Lending Club was listed on the NYSE, becoming the largest technology stock IPO of the year.
 - In 2015, the new loan amount for the Lending Club platform reached US\$8.36 billion.



Lending Club
IPO in 2014
Created in 2006

- Lending Club is a platform intermediary service that provides P2P loans to the market:
 - In December 2014, Lending Club was listed on the NYSE, becoming the largest technology stock IPO of the year.
 - In 2015, the new loan amount for the Lending Club platform reached US\$8.36 billion.
 - In the first half of 2016, Lending Club broke the scandal of illegal lending, and the founder left.

Outline

- 1 Data Selection
- 2 Data Exhibition**
- 3 Data Analysis
- 4 Future Work

- Ten features
 - id,gender,isVerified,year,month,day,province, follow_count, follower_count, tweet_count

- Ten features
 - `id,gender,isVerified,year,month,day,province,follow_count,follower_count,tweet_count`
- Each line shows one Weibo user's information

- Ten features
 - id,gender,isVerified,year,month,day,province, follow_count, follower_count, tweet_count
- Each line shows one Weibo user's information

```
1002698377,1,0,2013,07,20,3,40,5,3
```

User ID:1002698377, Male, Not Verified, 2013.07.20 registered,
Province: Shanxi, following 40 users, having 5 followers, posted 3 on Weibo

- Twelve features
 - `user_id`, `tweet_no`, `year`, `month`, `day`, `hour`, `minute`, `tweet_from`, `isForwarded`, `like_count`, `forward_count`, `review_count`

- Twelve features
 - `user_id,tweet_no,year,month,day,hour,minute,tweet_from, isForwarded,like_count,forward_count,review_count`
- Each line shows one Weibo user's information

- Twelve features
 - `user_id`, `tweet_no`, `year`, `month`, `day`, `hour`, `minute`, `tweet_from`, `isForwarded`, `like_count`, `forward_count`, `review_count`
- Each line shows one Weibo user's information

```
1002698377,1,2013,07,20,22,02,1,0,0,3,4
```

User ID:1002698377, this is the first Weibo that he/she
post:2013.07.20.22.02, from mobile device , receiving 0 like, 3
forwarded, 4 comments

- Fifty-seven features
 - Information of Applying for Loan, Information of Users, Information of Public Records, URL of Data

- Fifty-seven features
 - Information of Applying for Loan, Information of Users, Information of Public Records, URL of Data
- Each line shows one Weibo user's information

- Fifty-seven features
 - Information of Applying for Loan, Information of Users, Information of Public Records, URL of Data
- Each line shows one Weibo user's information

Solving the Missing Value

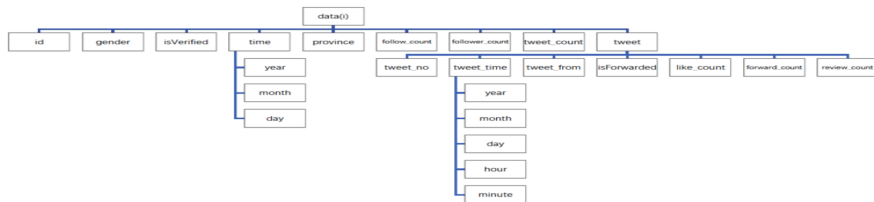
1. Deleting the according features
2. Filling the missing data

Outline

- 1 Data Selection
- 2 Data Exhibition
- 3 Data Analysis**
- 4 Future Work

Weibo Data

```
for i = 1 : size(data_1,1)
    data(i).id = data_1(i,1);
    data(i).gender = data_1(i, 2);
    data(i).isVerified = data_1(i, 3);
    data(i).time.year = data_1(i, 4);
    data(i).time.month = data_1(i, 5);
    data(i).time.day = data_1(i, 6);
    data(i).province = data_1(i, 7);
```



Lending Club Data

id	887379	non-null	int64
member_id	887379	non-null	int64
loan_amnt	887379	non-null	float64
funded_amnt	887379	non-null	float64
funded_amnt_inv	887379	non-null	float64
term	887379	non-null	object
int_rate	887379	non-null	float64
installment	887379	non-null	float64
grade	887379	non-null	object
sub_grade	887379	non-null	object
emp_title	835917	non-null	object
emp_length	842554	non-null	object
home_ownership	887379	non-null	object
annual_inc	887375	non-null	float64
verification_status	887379	non-null	object
issue_d	887379	non-null	object
loan_status	887379	non-null	object
pymnt_plan	887379	non-null	object
url	887379	non-null	object
desc	126028	non-null	object
purpose	887379	non-null	object
title	887227	non-null	object

Lending Club Data

zip_code	887379	non-null	object
addr_state	887379	non-null	object
dti	887379	non-null	float64
delinq_2yrs	887350	non-null	float64
earliest_cr_line	887350	non-null	object
inq_last_6mths	887350	non-null	float64
mths_since_last_delinq	433067	non-null	float64
mths_since_last_record	137053	non-null	float64
open_acc	887350	non-null	float64
pub_rec	887350	non-null	float64
revol_bal	887379	non-null	float64
revol_util	886877	non-null	float64
total_acc	887350	non-null	float64
initial_list_status	887379	non-null	object
out_prncp	887379	non-null	float64
out_prncp_inv	887379	non-null	float64
...			

Outline

- 1 Data Selection
- 2 Data Exhibition
- 3 Data Analysis
- 4 Future Work**

Future Work



Midterm Aim:
Visualization and further
analysis
Final Aim:
Designing Algorithm

– Yi Cui