

Sentiment Analysis on Tweets



Welcome to
Twitter

See what's happening in the world
right now

Sign up

Yuan Ying - ying.yua@husky.neu.edu

Mushtaq Rizvi - rizvi.m@husky.neu.edu

Wei Huang - huang.wei3@husky.neu.edu

JinJin Zhang - zhang.jinj@husky.neu.edu



PROJECT GOAL

- PROCESS REAL TWITTER DATASETS
- UTILIZE TWITTER API
- EXTRACT MEANINGFUL DATA
- NLP(NATURAL LANGUAGE PROCESS)
- PERFORM SENTIMENT ANALYSIS

USER INPUT



- A CITY NAME
- OR
- A COMPANY NAME

SYSTEM OUTPUT



- SENTIMENT SCORE FOR
- THE CITY WEATHER IN LAST 7 DAYS
- OR
- THE COMPANY STOCK IN LAST 7 DAYS

USER INPUT



- A KEYWORD
OR
- NOTHING

SYSTEM OUTPUT



- TOP 10 POPULAR HASHTAGS FROM UPCOMING TWITTER CONTAINING THAT KEYWORD IN A 5MIN WINDOW
- THE COUNT OF TWEETS AND SENTIMENT SCORE FOR EACH HASHTAG
- A BAR CHART SHOWING THE RESULT*

USER INPUT



- A KEYWORD
OR
- NOTHING

SYSTEM OUTPUT



- TOP 10 POPULAR LOCATIONS FROM UPCOMING TWITTER CONTAINING THAT KEYWORD IN A 10HOUR WINDOW
- THE COUNT OF TWEETS AND SENTIMENT SCORE FOR EACH LOCATION
- A MAP SHOWING THE RESULT*

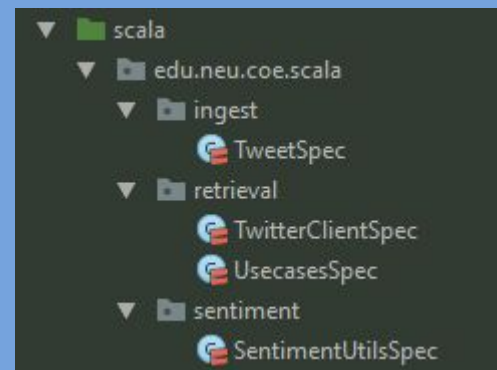
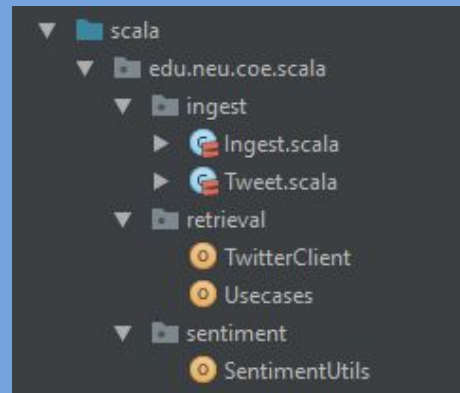


METH- ODOLOGY

- ACQUIRING (SEARCH API & STREAMING API)
- PARSING (JSON FORMAT)
- FILTERING (LANGUAGE & GEOLOCATION)
- CLEANING (SPECIAL CHARACTERS)
- CALCULATING SENTIMENT (STANFORD NLP)
- MAPPING & REDUCING (SPARK)
- VISUALIZATION (APACHE ZEPPELIN)

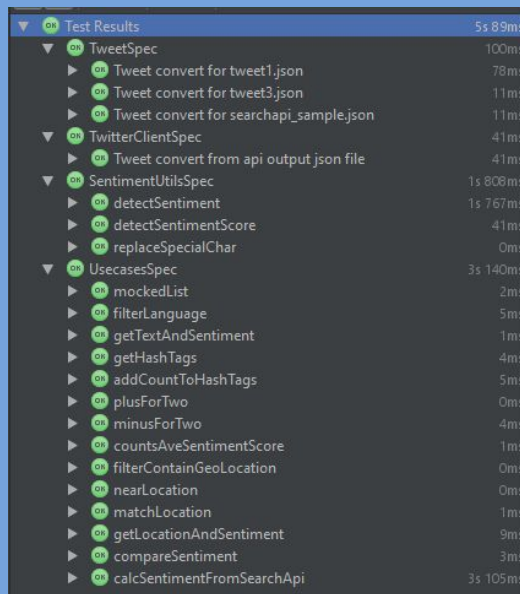
PROGRAMMING

- Ingest.scala
- Tweet.scala
- TwitterClient
- Usecases
- SentimentUtils
- Unit Test



UNIT TEST




- TOTAL 43 UNIT TESTS
- 74% CLASSES COVERED
- 60% LINES COVERED



Test Results	5s 89ms
▼ TweetSpec	100ms
▶ Tweet convert for tweet1.json	78ms
▶ Tweet convert for tweet3.json	11ms
▶ Tweet convert for searchapi_sample.json	11ms
▼ TwitterClientSpec	41ms
▶ Tweet convert from api output json file	41ms
▼ SentimentUtilsSpec	1s 808ms
▶ detectSentiment	1s 767ms
▶ detectSentimentScore	41ms
▶ replaceSpecialChar	0ms
▼ UsecasesSpec	3s 140ms
▶ mockedList	2ms
▶ filterLanguage	5ms
▶ getTextAndSentiment	1ms
▶ getHashTags	4ms
▶ addCountToHashTags	5ms
▶ plusForTwo	0ms
▶ minusForTwo	4ms
▶ countsAveSentimentScore	1ms
▶ filterContainGeoLocation	0ms
▶ nearLocation	0ms
▶ matchLocation	1ms
▶ getLocationAndSentiment	9ms
▶ compareSentiment	3ms
▶ calcSentimentFromSearchApi	3s 105ms

ACCEPTANCE CRITERIA



- VERIFY ANALYSIS RESULT WITH TEST TWEETS
 - Included in unit test
 - The accuracy should reach 90% 
- VERIFY ANALYSIS RESULT WITH KEYWORD RELATED TO WEATHER JUDGEMENT
 - Compare 2 city's weather in 7 days
 - The accuracy should reach 70% 
- VERIFY ANALYSIS RESULT WITH KEYWORD RELATED TO STOCK ASSESSMENT
 - Compare 2 company's stock in 7 days
 - The accuracy should reach 80% 



WEATHER

Boston vs. New York



90 per day per city(630 per city)

BOSTON GRADE: 1.35555556

NEW YORK GRADE: 1.41587301

PREDICTION: New York better than Boston

REAL WEATHER REPORT:

 56° 40°	 52° 45°	 57° 38°	 60° 47°	 55° 45°	 52° 44°	 57° 43°
 62° 42°	 73° 51°	 75° 58°	 75° 55°	 64° 48°	 63° 50°	 61° 48°

TRUTH: New York more pleasant than Boston

RESULT: accurate



WEATHER

Boston vs. SF



90 per day per city(630 per city)

BOSTON GRADE: 1.35555556

SF GRADE: 1.20634920

PREDICTION: Boston better than San Fran

REAL WEATHER REPORT:

 56° 40°	 52° 45°	 57° 38°	 60° 47°	 55° 45°	 52° 44°	 57° 43°
 59° 44°	 62° 47°	 64° 53°	 64° 53°	 58° 50°	 60° 47°	 66° 48°

TRUTH: San Francisco more pleasant than Boston

RESULT: in-accurate



WEATHER ANALYSIS SUMMARY

Among 10 cities over 7 days

TOTAL PREDICTION ACCURACY:

8/10 = 80%

SUCCESSFULLY ACHIEVED

① ACCEPTANCE CRITERIA reach 80%



APPLE INC. GRADE: 1.09797297
GNC GRADE : 1.06976744
PREDICTION: Apple Inc. better than GNC

REAL STOCK REPORT:

DATE	OPEN	HIGH	LOW	CLOSE	
04/17/2017	141.48	141.88	140.87	141.83	16,529,130
04/13/2017	141.91	142.38	141.05	141.05	17,775,510
04/12/2017	141.6	142.15	141.01	141.8	20,320,420
04/11/2017	142.94	143.35	140.06	141.63	30,341,520

04/17/2017	7.31	7.38	7.08	7.24	2,968,706
04/13/2017	7.26	7.33	7.015	7.3	2,624,080
04/12/2017	7.44	7.44	7.21	7.27	2,404,252
04/11/2017	7.14	7.51	7.07	7.42	3,792,926

TRUTH: GNC better than Apple Inc.
RESULT: in-accurate



STOCK

Apple Inc. vs. Netflix



90 per day per company(630 per company)

APPLE INC. GRADE: 1.09797297

Netflix GRADE : 1.41774193

PREDICTION: Netflix better than Apple Inc.

REAL STOCK REPORT:

DATE	OPEN	HIGH	LOW	CLOSE	
04/17/2017	141.48	141.88	140.87	141.83	16,529,130
04/13/2017	141.91	142.38	141.05	141.05	17,775,510
04/12/2017	141.6	142.15	141.01	141.8	20,320,420
04/11/2017	142.94	143.35	140.06	141.63	30,341,520
04/17/2017	144.43	147.32	144.43	147.25	15,993,460
04/13/2017	144.25	144.55	142.76	142.92	3,676,127
04/12/2017	144.85	145.74	143.55	143.83	4,383,524
04/11/2017	144.28	144.54	141.98	144.35	4,619,263

TRUTH: Netflix better than Apple Inc.

RESULT: accurate



STOCK ANALYSIS SUMMARY

Among 10 cities over 7 days

TOTAL PREDICTION ACCURACY:

$6/10 = 60\%$

FAILED

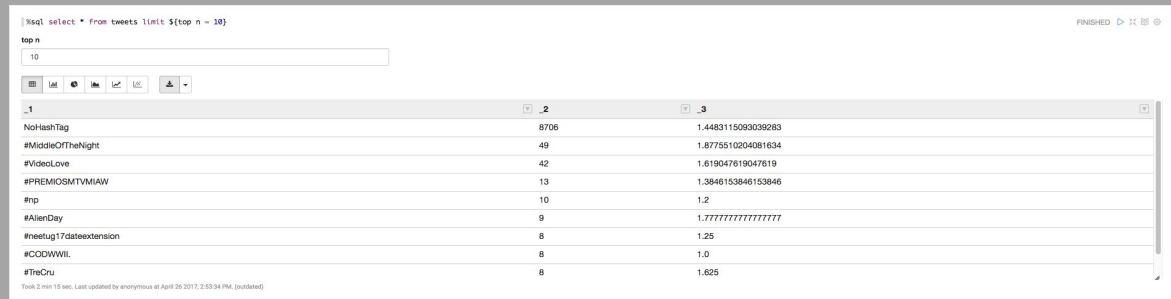


ACCEPTANCE CRITERIA reach 80%

MOST POPULAR HASHTAG

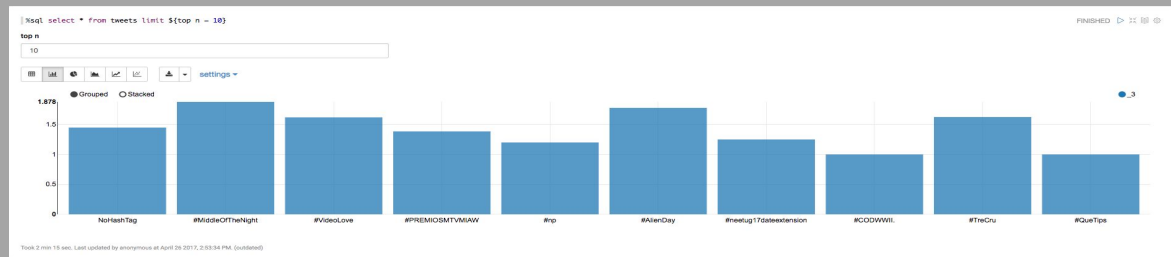
5mins duration

TABLE:



1	2	3
No!ashTag	8706	1.4483115093039283
#MiddleOfTheNight	49	1.8775510204081634
#VideoLove	42	1.619047619047619
#PREMIOSMTVMIAW	13	1.3846153846153846
#np	10	1.2
#AlienDay	9	1.7777777777777777
#meeting17dateextension	8	1.25
#CODWWII	8	1.0
#TheCru	8	1.625

BAR CHART:



TOTAL DIFFERENT HASHTAG COUNT: 2279

TOTAL TWEETS PROCESSED: 16896



SENTIMENT MAPPING

10 hour duration

CODE REPOSITORY



https://github.com/yingy4/CSYE7200_FinalProject_Team2_Spring2017

REFERENCE

- Course Repo
- <http://140dev.com/twitter-api-programming-tutorials/aggregating-tweets-search-api-vs-streaming-api/>
- <https://www.udemy.com/apache-spark-with-scala-hands-on-with-big-data/learn/v4/overview>
- <https://community.hortonworks.com/content/kbentry/90320/add-leaflet-map-to-zeppelin-notebook.html>
- <https://dev.twitter.com/rest/public>
- <https://github.com/vspiewak/twitter-sentiment-analysis/blob/master/src/main/scala/com/github/vspiewak/util/SentimentAnalysisUtils.scala>

THANK YOU!