analysis of job openings on twitter

Joyce Duan

Motivation - Why Twitter

Opportunities:

- Twitter provides real-time information about job openings
- Search for relevant hashtags (i.e.) #jobsearch produces many irrelevant results, i.e. articles on job hunting, random thoughts on jobs, and junk tweets that use the hashtag for promotion

Solution:

- Use machine learning to search for and curate job openings on twitter

Results Summary

- Collected and identified tweets about job openings
  - classification model: accuracy 92%, area under the curve: 0.96

- Data insights about job openings on twitter:
  - Most active hour of the day (PDT): 9am, 12pm - 1pm, 3pm
  - Regions with large number of job opening tweets:
    - Boston: large # of openings in most job categories
    - San Francisco and LA: jobs in marketing/sales
    - Austin: engineering openings
    - Houston: medical related openings

caveats: preliminary results. data insights were based on sampling of job opening tweets posted on 8/13 from 8am - 9pm PDT. Analysis of a larger dataset will validate if the observations are general trends.

Stats

time range: Aug 13th 8am - 8pm, sample 10 minutes per hour

tweets collected: 65,950

of job opening tweets: 47,853

unique screen_name: 10,259

unique user_locations: 2,786
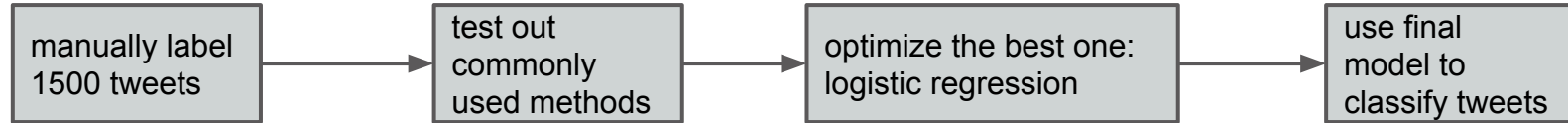
Future Directions

- Refine classification model and additional data cleaning
  - larger training set; additional features: i.e. valid url; info about twitter handle
- Extract and standardize info about jobs: standardized location, job title
- How to use this information:
  - Identify regions with supply shortage for certain job categories: i.e. IT
  - Generate candidate leads for companies with openings
  - Intelligent job alert agent for job seekers matching their resume to job description
- Other recruiting related valuable data on twitter: i.e. companies layoff, M&A, restructuring

  example tweets: #BREAKING Defence contractor BAE Systems has confirmed speculation it's cutting 125 jobs from it's Williamstown naval shipyard

Analysis Process

-Twitter Stream: listen for 10 minutes every hour

     stream.filter(track=['jobs', 'hiring', 'job'], languages=['en'])

- Filter: hashtag contains 'job', 'hire' 'hiring' or url contains: 'career', 'job'
- Predictive modeling: classify if tweets are about job openings or not
- Filter out spam and dedupe
- Data insights

Classification - work flow

```
manually label          test out              optimize the best one:      use final
1500 tweets             commonly               logistic regression         model to
                        used methods                                       classify tweets
```

5-fold cross validation
grid search on hyperparameters
scoring = accuracy

training set: (1125, 4847)
model, accuracy, precision, recall, f1
LogisticRegression ['0.909', '0.896', '0.976', '0.934']
     MultinomialNB ['0.883', '0.859', '0.984', '0.917']
     SVC ['0.661', '0.661', '1.000', '0.796']
RandomForest['0.907', '0.902', '0.964', '0.932']
     AdaBoostClassifier ['0.880', '0.911', '0.907', '0.909']
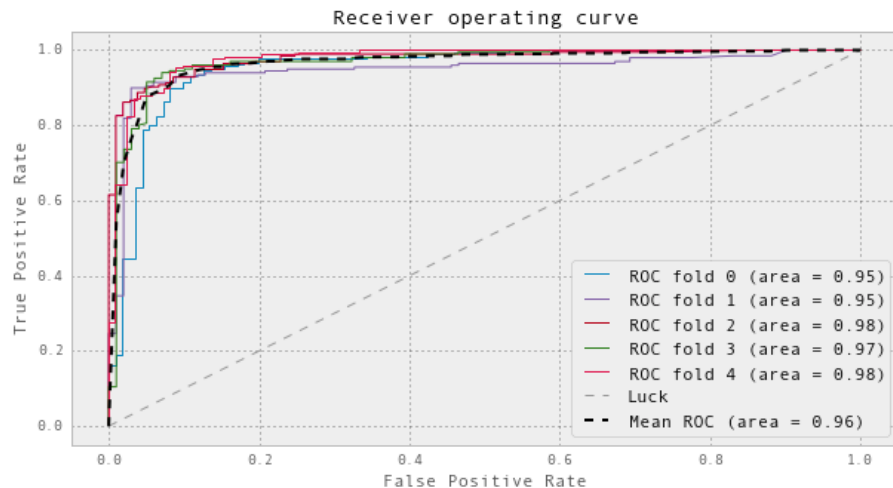##   KNeighborsClassifier ['0.904', '0.914', '0.944', '0.929']
     SVC ['0.661', '0.661', '1.000', '0.796']
     SVC ['0.872', '0.852', '0.976', '0.910']
     finish in 0.0083min for SVC

Classification - Final Model: logistic regression

Receiver operating curve

ROC fold 0 (area = 0.95)
ROC fold 1 (area = 0.95)
ROC fold 2 (area = 0.98)
ROC fold 3 (area = 0.97)
ROC fold 4 (area = 0.98)
Luck
Mean ROC (area = 0.96)

words signalling tweets are about job openings
sized by coefficients of the regression model

manager
hiring
associate usa
sales in apply
engineer
opening
specialist now ca jobs
assistant
details
looking

Data Insights

Most active hour of the day: 9am, 12pm - 1pm, 3pm PST
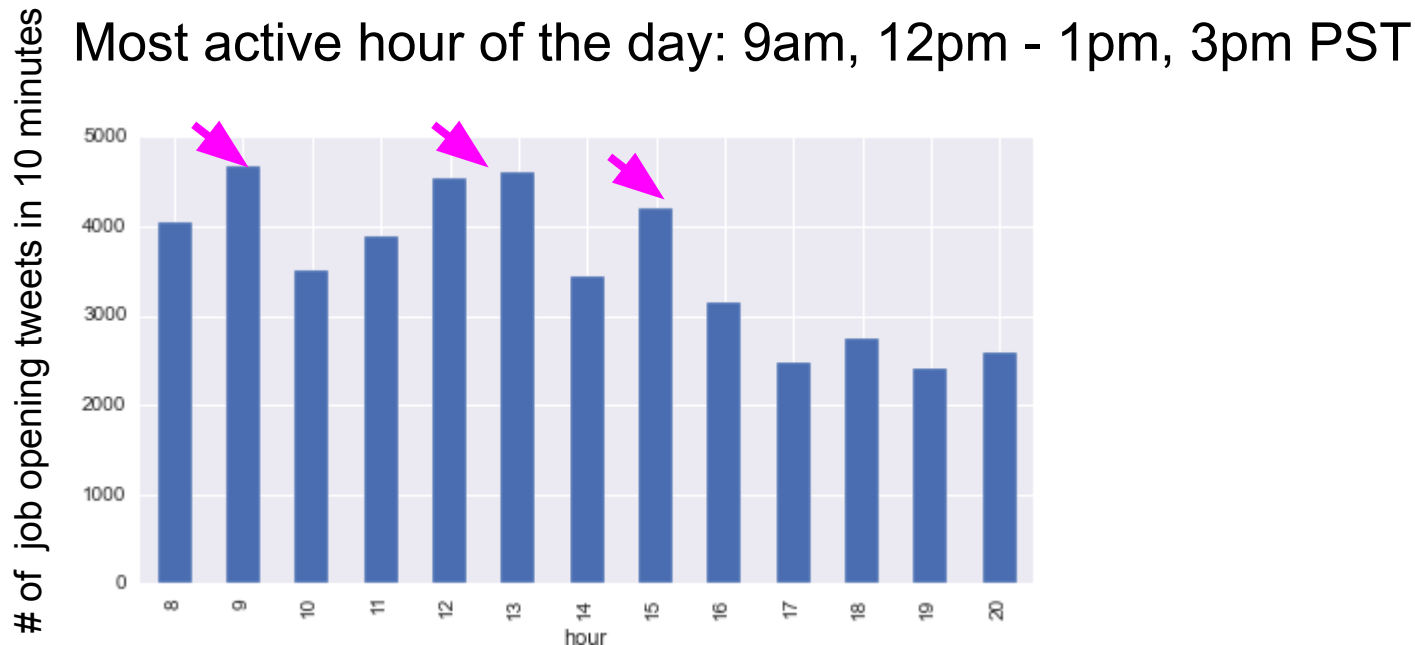
Data Insights - geo & job categories

	admin	analysis	driver	engineer	marketing/sales	medical
unknown	114	179	120	575	553	351
austin texas	6	21	0	67	13	2
boston ma	17	31	2	50	44	43
bristol england	19	1	6	12	7	13
california usa	2	3	34	28	4	0
columbus ohio	2	33	0	37	6	0
houston tx	6	5	2	14	33	56
leeds england	19	6	9	14	5	12
los angeles ca	7	4	1	9	64	44
michigan	6	25	34	39	9	8
san francisco ca	9	20	1	35	64	29
texas	6	6	11	23	19	45
united kingdom	21	7	7	22	16	11
usa	32	47	18	162	124	24
wisconsin	0	1	41	1	11	12
worldwide	12	13	1	100	21	6
Total	1424	2100	2039	5686	4452	2963

- Boston has large # of openings in most categories
- San Francisco and LA have jobs in marketing/sales
- Austin has many engineering openings
- Houston: many medical related openings

- preliminary results: location and job categories needs further cleaning
- user_location were used as surrogate; ~12% no location info
- counts represents relative number of openings on twitter; which is affected by actual number of openings and prevalence of twitter usage.

Hashtags usage

~10% of tweets had no hashtag;

~50% had 3-4 hashtags
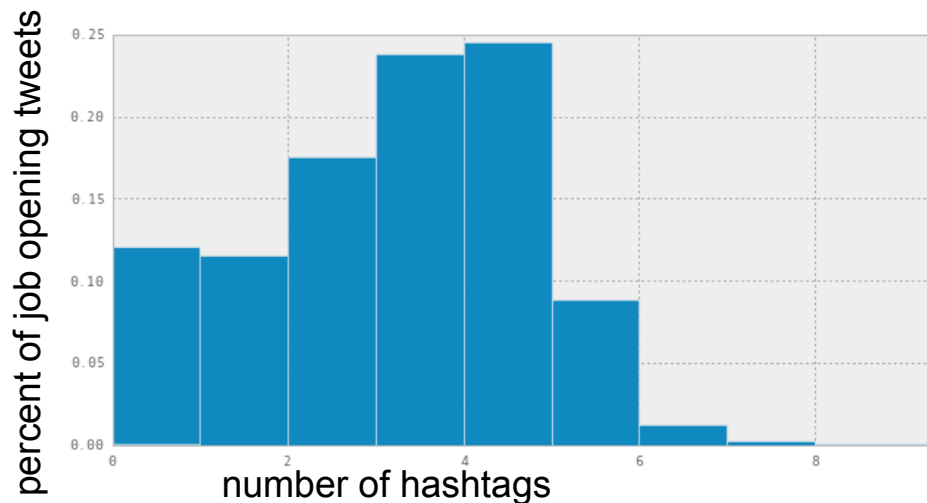
Spam

- large number of tweets from the same user
- same message tweeted up to 20 times in a single day , career sites from different screen_name

index	5174	5184	5204	5210	5221
created_at	Wed Aug 12 16:15:05 +0000 2015	Wed Aug 12 16:15:05 +0000 2015	Wed Aug 12 16:15:06 +0000 2015	Wed Aug 12 16:15:06 +0000 2015	Wed Aug 12 2015
created_by	Transport_work	Job_trucking	Jobs_transport_	jobs_trucking	Recruiter_tra
favorite_count	0	0	0	0	0
hash_tags	truck,hiring,job,Woodland,jobs	truck,hiring,job,Woodland,jobs	truck,hiring,job,Woodland,jobs	truck,hiring,job,Woodland,jobs	truck,hiring,j
id_str	631499154800312320	631499155370737665	631499157505622016	631499158243819520	6314991590
in_reply_to	None	None	None	None	None
retweet_count	0	0	0	0	0
text	★JOB★ #truck #hiring #job #Woodland - Company ...	★JOB★ #truck #hiring #job #Woodland - Company ...	★JOB★ #truck #hiring #job #Woodland - Company ...	★JOB★ #truck #hiring #job #Woodland - Company ...	★JOB★ #tr #Woodland
urls	http://j.obs.link/BqTR4f	http://j.obs.link/BqTSXH	http://j.obs.link/BqTSH3	http://j.obs.link/BqTQMY	http://j.obs.l
str_text	job #truck #hiring #job #woodland - company dr...	job #truck #hiring #job #woodland - company dr...	job #truck #hiring #job #woodland - company dr...	job #truck #hiring #job #woodland - company dr...	job #truck #l #woodland -

References

- time article: How twitter can become the premier site for job-seekers
- report: how companies engage talents on twitter