Report: Cmpe 493 Text Classification using Naive Bayes

Selim Öztürk 2013401027 April 5 2018 00:53

Number of Docs

Training set {'crude': 361, 'acq': 1617, 'grain': 429, 'earn': 2848, 'money-fx': 536} **Test set** {'crude': 152, 'acq': 638, 'grain': 139, 'earn': 933, 'money-fx': 149}

The 50 Most Informative Words

crude

refineri qtr ga gasolin share pipelin opec api explor arabia wheat energi inc gulf field profit reuter petroleum note dlr 3 crude net ct venezuela ecuador product oil iran minist cubic price quota dai refin dividend bank loss herrington bpd iranian mln sale record saudi drill output barrel tanker petrobra

acq

control own qtr cyclop sharehold share tonn acquisit rate purchas export file year tender inc sell said group acquir record merger twa stake to outstand net ct approv complet firm purol oil takeov hold offer dollar bid gencorp bui corp compani ha usair loss shr secur mln profit common stock

grain

oper certif depart usda 092 share crop tonn vs export soybean feed inc subsidi barlei area profit agricultur maiz dlr program farmer import net ct oil wheat farm acreag corn 198687 rain corp ec compani harvest bank winter loss rice lyng mln eep season acr weather grain commod soviet bushel

earn

oper qtr offici dollar tonn vs netshr at year avg market ct 1986 said wa would to profit offer rev note dlr split net record lossshr oil thei that price earn agreement 4th gain not dividend ha by loss shr 31 mln exclud prior mth bui grain export div quarter

money-fx

bill england yen exchang dollar trade rate at monei reserv japan economist dealer market ct fed currenc profit dlr that econom economi to deficit bundesbank net inc germani around stabil share pari band corp shortag compani monetari bank loss stg central interven treasuri mln foreign polici uk further intervent sterl

Summary

alpha = 0

using the 50 most informative words

	precision	recall	f-score
micro-avgs	0.82	0.82	0.82
macro-avgs	0.89	0.66	0.76

using all words

	precision	recall	f-score
micro-avgs	0.91	0.91	0.91
macro-avgs	0.95	0.88	0.91

alpha = 1

using the 50 most informative words

	precision	recall	f-score
micro-avgs	0.82	0.82	0.82
macro-avgs	0.91	0.64	0.75

using all words

	precision	recall	f-score
micro-avgs	0.92	0.92	0.92
macro-avgs	0.96	0.82	0.89

Screenshots

```
reading file Dataset/reut2-010.sgm
reading file Dataset/reut2-020.sgm
docs_in_test_set 2011
docs_in_training_set 5791
test vocabulary length 16436
training vocabulary length 31163
```

using the 50 most informative words alpha = 0	using the EQ most	informative year	1-	
recall f-score crude 0.83 0.32 0.46 acq 0.93 0.83 0.88 grain 0.98 0.39 0.56 earn 0.75 0.97 0.84 money-fx 0.94 0.79 0.86 micro-avgs 0.82 0.82 0.82 macro-avgs 0.89 0.66 0.76 using all words alpha = 0 precision recall f-score crude 0.98 0.73 0.84 acq 0.93 0.86 0.89 grain 0.99 0.88 0.93 grain 0.99 0.88 0.93 money-fx 0.99 0.93 0.96 micro-avgs 0.91 0.91 0.91 macro-avgs 0.95 0.88 0.91 using the 50 most informative words alpha = 1 precision recall f-score crude 0.98 0.73 0.86 0.89 grain 0.99 0.88 0.93 0.96 micro-avgs 0.91 0.91 0.91 macro-avgs 0.95 0.88 0.91 using the 50 most informative words alpha = 1 precision recall f-score crude 0.95 0.23 0.37 acq 0.92 0.84 0.88 grain 0.98 0.40 0.57 earn 0.75 0.97 0.84 money-fx 0.96 0.76 0.85 micro-avgs 0.82 0.82 0.82 macro-avgs 0.91 0.64 0.75 using all words alpha = 1 precision recall f-score crude 1.00 0.53 0.70 acq 0.88 0.96 0.92 grain 1.00 0.53 0.70 acq 0.88 0.96 0.92 grain 1.00 0.73 0.85 earn 0.92 0.99 0.96 money-fx 1.00 0.91 0.95 micro-avgs 0.92 0.99 0.96 money-fx 1.00 0.91 0.95 micro-avgs 0.92 0.99 0.96		IIIIOIMACIVE WOLU	15	
crude 0.83 0.32 0.46 acq 0.93 0.83 0.88 grain 0.98 0.39 0.56 earn 0.75 0.97 0.84 money-fx 0.94 0.79 0.86 micro-avgs 0.82 0.82 0.82 micro-avgs 0.89 0.66 0.76 using all words alpha = 0 Precision recall f-score crude 0.98 0.73 0.84 0.84 acq 0.93 0.86 0.89 grain 0.99 0.88 0.93 money-fx 0.99 0.93 0.96 micro-avgs 0.91 0.91 0.91 using the 50 most informative words alpha = 1 precision recall f-score crude 0.95 0.23 0.37 acq 0.92 0.84 0.88 grain 0.98 0.40 0.57 earn 0.75	атрпа = 0	neneirian	eseal1	£
acq 0.93 0.83 0.88 grain 0.98 0.39 0.56 earn 0.75 0.97 0.84 money-fx 0.94 0.79 0.86 micro-avgs 0.82 0.82 0.82 macro-avgs 0.89 0.66 0.76 using all words alpha = 0 precision recall f-score crude 0.98 0.73 0.84 acq 0.93 0.86 0.99 grain 0.99 0.88 0.94 earn 0.88 0.98 0.93 money-fx 0.99 0.93 0.96 micro-avgs 0.91 0.91 0.91 macro-avgs 0.95 0.88 0.91 using the 50 most informative words alpha = 1 precision recall f-score crude 0.95 0.23 0.37 acq 0.92 0.84 0.88 grain 0.98 0.40 0.57 earn 0.75 0.97 0.84 money-fx 0.96 0.76 0.85 micro-avgs 0.82 0.82 0.82 macro-avgs 0.91 0.64 0.75 using all words alpha = 1 precision recall f-score crude 0.95 0.23 0.37 acq 0.92 0.84 0.88 grain 0.98 0.40 0.57 earn 0.75 0.97 0.84 money-fx 0.96 0.76 0.85 micro-avgs 0.82 0.82 0.82 macro-avgs 0.91 0.64 0.75 using all words alpha = 1 precision recall f-score crude 1.00 0.53 0.70 acq 0.88 0.96 0.92 grain 1.00 0.53 0.70 acq 0.88 0.96 0.92 grain 1.00 0.73 0.85 earn 0.92 0.99 0.96 money-fx 1.00 0.91 0.95 micro-avgs 0.92 0.99	a ruida			
grain 0.98 0.39 0.56 earn 0.75 0.97 0.84 money-fx 0.94 0.79 0.86 micro-avgs 0.82 0.82 0.82 macro-avgs 0.89 0.66 0.76 using all words alpha = 0 precision recall f-score crude 0.98 0.73 0.86 earn 0.88 0.98 grain 0.99 0.88 0.93 money-fx 0.99 0.93 0.96 micro-avgs 0.91 0.91 0.91 macro-avgs 0.95 0.88 0.91 using the 50 most informative words alpha = 1 precision recall f-score crude 0.95 0.23 0.37 acq 0.92 0.84 0.88 grain 0.99 0.88 0.91 using the 50 most informative words alpha = 1 precision recall f-score crude 0.95 0.23 0.37 acq 0.92 0.84 0.88 grain 0.98 0.40 0.57 earn 0.75 0.97 0.84 money-fx 0.96 0.76 0.85 micro-avgs 0.91 0.64 0.75 using all words alpha = 1 precision recall f-score crude 1.00 0.53 0.70 acq 0.88 0.90 0.92 grain 1.00 0.53 0.70 acq 0.88 0.96 0.92 grain 1.00 0.73 0.85 earn 0.92 0.99 money-fx 1.00 0.91 0.95 micro-avgs 0.92 0.92	10000000			
earn 0.75 0.97 0.84 money-fx 0.94 0.79 0.86 micro-avgs 0.82 0.82 0.82 macro-avgs 0.89 0.66 0.76 using all words alpha = 0				
money-fx 0.94 0.79 0.86 micro-avgs 0.82 0.82 0.82 using all words alpha = 0 recision recall f-score crude 0.98 0.73 0.84 acq 0.93 0.86 0.89 grain 0.99 0.88 0.94 earn 0.88 0.98 0.93 money-fx 0.99 0.93 0.96 micro-avgs 0.91 0.91 0.91 macro-avgs 0.95 0.88 0.91 using the 50 most informative words alpha = 1 precision recall f-score crude 0.95 0.23 0.37 acq 0.92 0.84 0.88 grain 0.98 0.40 0.57 earn 0.75 0.97 0.84 money-fx 0.96 0.76 0.85 micro-avgs 0.91 0.64 0.75 using all words alpha = 1 precision				
micro-avgs 0.82 0.82 0.82 using all words alpha = 0				
macro-avgs 0.89 0.66 0.76 using all words alpha = 0	State of the state			
using all words alpha = 0 precision recall f-score crude 0.98 0.73 0.84 acq 0.93 0.86 0.89 grain 0.99 0.88 0.94 earn 0.88 0.98 0.93 money-fx 0.99 0.93 0.96 micro-avgs 0.91 0.91 0.91 using the 50 most informative words alpha = 1 precision recall f-score crude 0.95 0.23 0.37 acq 0.92 0.84 0.88 grain 0.98 0.40 0.57 earn 0.75 0.97 0.84 money-fx 0.96 0.76 0.85 micro-avgs 0.82 0.82 0.82 macro-avgs 0.91 0.64 0.75 using all words alpha = 1 precision recall f-score crude 1.00 0.53 0.70 acq 0.88 0.90 0.92 grain 1.00 0.73 0.85 earn 0.92 0.99 0.96 money-fx 1.00 0.91 0.95 micro-avgs 0.92 0.99				
alpha = 0		0.89	0.66	0.76
crude 0.98 0.73 0.84 acq 0.93 0.86 0.89 grain 0.99 0.88 0.94 earn 0.88 0.98 0.93 money-fx 0.99 0.93 0.96 micro-avgs 0.91 0.91 0.91 macro-avgs 0.95 0.88 0.91 using the 50 most informative words alpha = 1 f-score crude 0.95 0.23 0.37 acq 0.92 0.84 0.88 grain 0.98 0.40 0.57 earn 0.75 0.97 0.84 money-fx 0.96 0.76 0.85 micro-avgs 0.82 0.82 0.82 macro-avgs 0.91 0.64 0.75 using all words alpha = 1 f-score crude 1.00 0.53 0.70 acq 0.88 0.96 0.92 grain 1.00	The state of the s			
crude 0.98 0.73 0.84 acq 0.93 0.86 0.89 grain 0.99 0.88 0.94 earn 0.88 0.98 0.93 money-fx 0.99 0.93 0.96 micro-avgs 0.91 0.91 0.91 macro-avgs 0.95 0.88 0.91 using the 50 most informative words alpha = 1 f-score crude 0.95 0.23 0.37 acq 0.92 0.84 0.88 grain 0.98 0.40 0.57 earn 0.75 0.97 0.84 money-fx 0.96 0.76 0.85 micro-avgs 0.82 0.82 0.82 macro-avgs 0.91 0.64 0.75 using all words alpha = 1 precision recall f-score crude 1.00 0.53 0.70 0.92 grain 1.00 0.53 0.70	alpha = 0			
acq 0.93 0.86 0.89 grain 0.99 0.88 0.94 earn 0.88 0.98 0.93 money-fx 0.99 0.93 0.96 micro-avgs 0.91 0.91 0.91 macro-avgs 0.95 0.88 0.91 using the 50 most informative words alpha = 1 precision recall f-score crude 0.95 0.23 0.37 acq 0.92 0.84 0.88 grain 0.98 0.40 0.57 earn 0.75 0.97 0.84 money-fx 0.96 0.76 0.85 micro-avgs 0.82 0.82 0.82 macro-avgs 0.91 0.64 0.75 using all words alpha = 1 precision recall f-score crude 1.00 0.53 0.70 acq 0.88 0.96 0.92 grain 1.00 0.53 0.70 acq 0.88 0.96 0.92 grain 1.00 0.73 0.85 earn 0.92 0.99 0.96 money-fx 1.00 0.91 0.95 micro-avgs 0.92 0.99	32			
grain 0.99 0.88 0.94 earn 0.88 0.98 0.93 money-fx 0.99 0.93 0.96 micro-avgs 0.91 0.91 0.91 macro-avgs 0.95 0.88 0.91 using the 50 most informative words alpha 0.91 using the 50 most informative words alpha 0.91 alpha = 1 precision recall f-score crude 0.95 0.23 0.37 acq 0.92 0.84 0.88 grain 0.98 0.40 0.57 earn 0.75 0.97 0.84 money-fx 0.96 0.76 0.85 micro-avgs 0.82 0.82 0.82 macro-avgs 0.91 0.64 0.75 using all words alpha = 1 precision recall f-score crude 1.00 0.53 0.70 acq 0.88 0.96 0.92 <	crude			
earn 0.88 0.98 0.93 money-fx 0.99 0.93 0.96 micro-avgs 0.91 0.91 0.91 macro-avgs 0.95 0.88 0.91 using the 50 most informative words alpha = 1	acq	0.93	0.86	0.89
money-fx 0.99 0.93 0.96 micro-avgs 0.91 0.91 0.91 macro-avgs 0.95 0.88 0.91 using the 50 most informative words alpha = 1 recision recall f-score crude 0.95 0.23 0.37 acq 0.92 0.84 0.88 grain 0.98 0.40 0.57 earn 0.75 0.97 0.84 money-fx 0.96 0.76 0.85 micro-avgs 0.82 0.82 0.82 macro-avgs 0.91 0.64 0.75 using all words alpha = 1 precision recall f-score crude 1.00 0.53 0.70 acq 0.88 0.96 0.92 grain 1.00 0.73 0.85 grain 1.00 0.91 0.96 money-fx 1.00	grain	0.99	0.88	0.94
micro-avgs 0.91 0.91 0.91 macro-avgs 0.95 0.88 0.91 using the 50 most informative words alpha = 1 recision recall f-score crude 0.95 0.23 0.37 acq 0.92 0.84 0.88 grain 0.98 0.40 0.57 earn 0.75 0.97 0.84 money-fx 0.96 0.76 0.85 micro-avgs 0.82 0.82 0.82 macro-avgs 0.91 0.64 0.75 using all words alpha = 1 recall f-score crude 1.00 0.53 0.70 acq 0.88 0.96 0.92 grain 1.00 0.73 0.85 earn 0.92 0.99 0.96 money-fx 1.00 0.91 0.95 micro-avgs 0.92 0.92 0.92 0.92	earn	0.88	0.98	0.93
macro-avgs 0.95 0.88 0.91 using the 50 most informative words alpha = 1 crude 0.95 0.23 0.37 acq 0.92 0.84 0.88 grain 0.98 0.40 0.57 earn 0.75 0.97 0.84 money-fx 0.96 0.76 0.85 micro-avgs 0.82 0.82 0.82 macro-avgs 0.91 0.64 0.75 using all words alpha = 1 precision recall f-score crude 1.00 0.53 0.70 acq 0.88 0.96 0.92 grain 1.00 0.73 0.85 earn 0.92 0.99 0.96 money-fx 1.00 0.91 0.95 micro-avgs 0.92 0.92 0.92	money-fx	0.99	0.93	0.96
using the 50 most informative words alpha = 1 precision recall f-score crude 0.95 0.23 0.37 acq 0.92 0.84 0.88 grain 0.98 0.40 0.57 earn 0.75 0.97 0.84 money-fx 0.96 0.76 0.85 micro-avgs 0.82 0.82 0.82 macro-avgs 0.91 0.64 0.75 using all words alpha = 1 precision recall f-score crude 1.00 0.53 0.70 acq 0.88 0.96 0.92 grain 1.00 0.73 0.85 earn 0.92 0.99 0.96 money-fx 1.00 0.91 0.95 micro-avgs 0.92 0.92 0.92	micro-avgs	0.91	0.91	0.91
alpha = 1 precision recall f-score crude 0.95 0.23 0.37 acq 0.92 0.84 0.88 grain 0.98 0.40 0.57 earn 0.75 0.97 0.84 money-fx 0.96 0.76 0.85 micro-avgs 0.82 0.82 0.82 macro-avgs 0.91 0.64 0.75 using all words alpha = 1 precision recall f-score crude 1.00 0.53 0.70 acq 0.88 0.96 0.92 grain 1.00 0.73 0.85 earn 0.92 0.99 0.96 money-fx 1.00 0.91 0.95 micro-avgs 0.92 0.92 0.92	macro-avgs	0.95	0.88	0.91
precision recall f-score crude 0.95 0.23 0.37 acq 0.92 0.84 0.88 grain 0.98 0.40 0.57 earn 0.75 0.97 0.84 money-fx 0.96 0.76 0.85 micro-avgs 0.82 0.82 0.82 macro-avgs 0.91 0.64 0.75 using all words alpha = 1 precision recall f-score crude 1.00 0.53 0.70 acq 0.88 0.96 0.92 grain 1.00 0.73 0.85 earn 0.92 0.99 0.96 money-fx 1.00 0.91 0.95 micro-avgs 0.92 0.92 0.92	using the 50 most	informative word	ls	
crude 0.95 0.23 0.37 acq 0.92 0.84 0.88 grain 0.98 0.40 0.57 earn 0.75 0.97 0.84 money-fx 0.96 0.76 0.85 micro-avgs 0.82 0.82 0.82 macro-avgs 0.91 0.64 0.75 using all words alpha = 1 precision recall f -score crude 1.00 0.53 0.70 acq 0.88 0.96 0.92 grain 1.00 0.73 0.85 earn 0.92 0.99 0.96 money-fx 1.00 0.91 0.95 micro-avgs 0.92 0.92 0.92	alpha = 1			
acq 0.92 0.84 0.88 grain 0.98 0.40 0.57 earn 0.75 0.97 0.84 money-fx 0.96 0.76 0.85 micro-avgs 0.82 0.82 0.82 macro-avgs 0.91 0.64 0.75 using all words alpha = 1 precision recall f-score crude 1.00 0.53 0.70 acq 0.88 0.96 0.92 grain 1.00 0.73 0.85 earn 0.92 0.99 0.96 money-fx 1.00 0.91 0.95 micro-avgs 0.92 0.92		precision	recall	f-score
grain 0.98 0.40 0.57 earn 0.75 0.97 0.84 money-fx 0.96 0.76 0.85 micro-avgs 0.82 0.82 0.82 macro-avgs 0.91 0.64 0.75 using all words alpha = 1 precision recall f-score crude 1.00 0.53 0.70 acq 0.88 0.96 0.92 grain 1.00 0.73 0.85 earn 0.92 0.99 0.96 money-fx 1.00 0.91 0.95 micro-avgs 0.92 0.92	crude	0.95	0.23	0.37
earn 0.75 0.97 0.84 money-fx 0.96 0.76 0.85 micro-avgs 0.82 0.82 0.82 macro-avgs 0.91 0.64 0.75 using all words alpha = 1 precision recall f-score crude 1.00 0.53 0.70 acq 0.88 0.96 0.92 grain 1.00 0.73 0.85 earn 0.92 0.99 0.96 money-fx 1.00 0.91 0.95 micro-avgs 0.92 0.92	acq	0.92	0.84	0.88
money-fx 0.96 0.76 0.85 micro-avgs 0.82 0.82 0.82 macro-avgs 0.91 0.64 0.75 using all words alpha = 1 precision recall f -score crude 1.00 0.53 0.70 acq 0.88 0.96 0.92 grain 1.00 0.73 0.85 earn 0.92 0.99 0.96 money-fx 1.00 0.91 0.95 micro-avgs 0.92 0.92 0.92	grain	0.98	0.40	0.57
micro-avgs 0.82 0.82 0.82 macro-avgs 0.91 0.64 0.75 using all words alpha = 1 precision recall f-score crude 1.00 0.53 0.70 acq 0.88 0.96 0.92 grain 1.00 0.73 0.85 earn 0.92 0.99 0.96 money-fx 1.00 0.91 0.95 micro-avgs 0.92 0.92	earn	0.75	0.97	0.84
macro-avgs 0.91 0.64 0.75 using all words alpha = 1 crude precision recall f-score crude 1.00 0.53 0.70 acq 0.88 0.96 0.92 grain 1.00 0.73 0.85 earn 0.92 0.99 0.96 money-fx 1.00 0.91 0.95 micro-avgs 0.92 0.92 0.92	money-fx	0.96	0.76	0.85
<pre>using all words alpha = 1</pre>	micro-avgs	0.82	0.82	0.82
alpha = 1 precision recall f-score crude 1.00 0.53 0.70 acq 0.88 0.96 0.92 grain 1.00 0.73 0.85 earn 0.92 0.99 0.96 money-fx 1.00 0.91 0.95 micro-avgs 0.92 0.92 0.92	macro-avgs	0.91	0.64	0.75
precision recall f-score crude 1.00 0.53 0.70 acq 0.88 0.96 0.92 grain 1.00 0.73 0.85 earn 0.92 0.99 0.96 money-fx 1.00 0.91 0.95 micro-avgs 0.92 0.92 0.92	using all words			
crude 1.00 0.53 0.70 acq 0.88 0.96 0.92 grain 1.00 0.73 0.85 earn 0.92 0.99 0.96 money-fx 1.00 0.91 0.95 micro-avgs 0.92 0.92 0.92	alpha = 1			
acq 0.88 0.96 0.92 grain 1.00 0.73 0.85 earn 0.92 0.99 0.96 money-fx 1.00 0.91 0.95 micro-avgs 0.92 0.92 0.92		precision	recall	f-score
grain 1.00 0.73 0.85 earn 0.92 0.99 0.96 money-fx 1.00 0.91 0.95 micro-avgs 0.92 0.92 0.92	crude	1.00	0.53	0.70
grain 1.00 0.73 0.85 earn 0.92 0.99 0.96 money-fx 1.00 0.91 0.95 micro-avgs 0.92 0.92 0.92	acq	0.88	0.96	0.92
earn 0.92 0.99 0.96 money-fx 1.00 0.91 0.95 micro-avgs 0.92 0.92 0.92	grain	1.00	0.73	0.85
money-fx 1.00 0.91 0.95 micro-avgs 0.92 0.92 0.92	198			
micro-avgs 0.92 0.92 0.92				
	macro-avgs	0.96	0.82	0.89