

INTRODUCTION TO WEB SCIENCES: Assignment 10

Babitha Bokka

11 December 2014

Contents

1	Question 1:	2
1.1	Approach	2
2	Question 2:	3
2.1	Approach	3
2.2	Source Code	4
2.2.1	docclass.py	4
2.2.2	feedfilter.py	8
2.2.3	entries.py	10
3	Question 3:	21
3.1	Approach	21

1 Question 1:

Choose a blog or a newsfeed (or something similar as long as it has an Atom or RSS feed). It should be on a topic or topics of which you are qualified to provide classification training data. In other words, choose something that you enjoy and are knowledgeable of. Find a feed with at least 100 entries.

Create between four and eight different categories for the entries in the feed:

examples:

work, class, family, news, deals

liberal, conservative, moderate, libertarian

sports, local, financial, national, international, entertainment

metal, electronic, ambient, folk, hip-hop, pop

Download and process the pages of the feed as per the week 12 class slides.

1.1 Approach

I started working on the assignment by choosing a blog
<http://orangette.blogspot.com/>.

1. The hardest part of the assignment is choosing a blog.
2. This is a blog which describes about a person important days and the type of food.
3. So, the categories I choose are
 - personal: When the author talks about her personal life.
 - cooking: About food and recipe.
 - place: Describes about places and traveling.
 - weather: Describes about the month, season and days.
4. Downloaded and saved the 100 entries in an xml file orangette.xml.

2 Question 2:

Manually classify the first 50 entries, and then classify (using the fisher classifier) the remaining 50 entries. Report the `cprob()` values for the 50 titles as well. From the title or entry itself, specify the 1-, 2-, or 3-gram that you used for the string to classify. Do not repeat strings; you will have 50 unique strings. For example, in these titles the string used is marked with `*s`:

`*Rachel Goswell*` - "Waves Are Universal" (LP Review)
The `*Naked and Famous*` - "Passive Me, Aggressive You" (LP Review)
`*Negativland*` - "Live at Lewis's, Norfolk VA, November 21, 1992" (concert)
Negativland - `"*U2*"` (LP Review)

Note how "Negativland" is not repeated as a classification string.

Create a table with the title, the string used for classification, `cprob()`, predicted category, and actual category.

2.1 Approach

1. To solve this question I started my approach by using `docclass.py` and `feedfilter.py`.
2. The first 50 entries are manually classified and other 50 entries are classified using the fisher method.
3. Table 1 gives you the manually classified the entries.
4. Table 2 gives you the classification based on fisher method.

2.2 Source Code

2.2.1 docclass.py

```
1 from pysqlite2 import dbapi2 as sqlite
2 import re
3 import math
4
5 def getwords(doc):
6     splitter=re.compile('\W*')
7     print doc
8     # Split the words by non-alpha characters
9     words=[s.lower() for s in splitter.split(doc)
10            if len(s)>2 and len(s)<20]
11
12     # Return the unique set of words only
13     return dict([(w,1) for w in words])
14
15 class classifier:
16     def __init__(self,getfeatures,filename=None):
17         # Counts of feature/category combinations
18         self.fc={}
19         # Counts of documents in each category
20         self.cc={}
21         self.getfeatures=getfeatures
22
23     def setdb(self,dbfile):
24         self.con=sqlite.connect(dbfile)
25         self.con.execute('create table if not exists fc(feature,category,count)')
26         self.con.execute('create table if not exists cc(category,count)')
27
28
29     def incf(self,f,cat):
30         count=self.fcount(f,cat)
31         if count==0:
32             self.con.execute("insert into fc values ('%s','%s',1)"
33                             "% (f,cat))"
34         else:
35             self.con.execute(
36                 "update fc set count=%d where feature='%s' and category='%s'"
37                 "% (count+1,f,cat))"
38
39     def fcount(self,f,cat):
40         res=self.con.execute(
41             'select count from fc where feature="%s" and category="%s"'
42             "%(f,cat)).fetchone()"
43         if res==None: return 0
44         else: return float(res[0])
45
46     def incc(self,cat):
47         count=self.catcount(cat)
48         if count==0:
49             self.con.execute("insert into cc values ('%s',1)" % (cat))
50         else:
51             self.con.execute("update cc set count=%d where category='%s'"
52                             "% (count+1,cat))"
53
```

```

54 def catcount(self,cat):
55     res=self.con.execute('select count from cc where category="%s" '
56                           %(cat)).fetchone()
57     if res==None: return 0
58     else: return float(res[0])
59
60 def categories(self):
61     cur=self.con.execute('select category from cc');
62     return [d[0] for d in cur]
63
64 def totalcount(self):
65     res=self.con.execute('select sum(count) from cc').fetchone();
66     if res==None: return 0
67     return res[0]
68
69
70 def train(self,item,cat):
71     features=self.getfeatures(item)
72     # Increment the count for every feature with this category
73     for f in features:
74         self.incf(f,cat)
75
76     # Increment the count for this category
77     self.incc(cat)
78     self.con.commit()
79
80 def fprob(self,f,cat):
81     if self.catcount(cat)==0: return 0
82
83     # The total number of times this feature appeared in this
84     # category divided by the total number of items in this category
85     return self.fcount(f,cat)/self.catcount(cat)
86
87 def weightedprob(self,f,cat,prf,weight=1.0,ap=0.5):
88     # Calculate current probability
89     basicprob=prf(f,cat)
90
91     # Count the number of times this feature has appeared in
92     # all categories
93     totals=sum([self.fcount(f,c) for c in self.categories()])
94
95     # Calculate the weighted average
96     bp=((weight*ap)+(totals*basicprob))/(weight+totals)
97     return bp
98
99
100
101
102 class naivebayes(classifier):
103
104     def __init__(self,getfeatures):
105         classifier.__init__(self,getfeatures)
106         self.thresholds={}
107
108     def docprob(self,item,cat):
109         features=self.getfeatures(item)
110

```

```

111 # Multiply the probabilities of all the features together
112 p=1
113 for f in features: p*=self.weightedprob(f,cat,self.fprob)
114 return p
115
116 def prob(self,item,cat):
117     catprob=self.catcount(cat)/self.totalcount()
118     docprob=self.docprob(item,cat)
119     return docprob*catprob
120
121 def setthreshold(self,cat,t):
122     self.thresholds[cat]=t
123
124 def getthreshold(self,cat):
125     if cat not in self.thresholds: return 1.0
126     return self.thresholds[cat]
127
128 def classify(self,item,default=None):
129     probs={}
130     # Find the category with the highest probability
131     max=0.0
132     for cat in self.categories():
133         probs[cat]=self.prob(item,cat)
134         if probs[cat]>max:
135             max=probs[cat]
136             best=cat
137
138     # Make sure the probability exceeds threshold*next best
139     for cat in probs:
140         if cat==best: continue
141         if probs[cat]*self.getthreshold(best)>probs[best]: return default
142     return best
143
144 class fisherclassifier(classifier):
145     def cprob(self,f,cat):
146         # The frequency of this feature in this category
147         clf=self.fprob(f,cat)
148         if clf==0: return 0
149
150         # The frequency of this feature in all the categories
151         freqsum=sum([self.fprob(f,c) for c in self.categories()])
152
153         # The probability is the frequency in this category divided by
154         # the overall frequency
155         p=clf/(freqsum)
156
157     return p
158     def fisherprob(self,item,cat):
159         # Multiply all the probabilities together
160         p=1
161         features=self.getfeatures(item)
162         for f in features:
163             p*=(self.weightedprob(f,cat,self.cprob))
164
165         # Take the natural log and multiply by -2
166         fscore=-2*math.log(p)
167

```

```

168     # Use the inverse chi2 function to get a probability
169     return self.invchi2(fscore, len(features)*2)
170 def invchi2(self, chi, df):
171     m = chi / 2.0
172     sum = term = math.exp(-m)
173     for i in range(1, df//2):
174         term *= m / i
175         sum += term
176     return min(sum, 1.0)
177 def __init__(self, getfeatures):
178     classifier.__init__(self, getfeatures)
179     self.minimums={}
180
181 def setminimum(self, cat, min):
182     self.minimums[cat]=min
183
184 def getminimum(self, cat):
185     if cat not in self.minimums: return 0
186     return self.minimums[cat]
187 def classify(self, item, default=None):
188     # Loop through looking for the best result
189     best=default
190     max=0.0
191     for c in self.categories():
192         p=self.fisherprob(item, c)
193         # Make sure it exceeds its minimum
194         if p>self.getminimum(c) and p>max:
195             best=c
196             max=p
197     return best

```


2.2.2 feedfilter.py

```
1 import feedparser
2 import re
3 import math
4 import docclass
5
6
7 # Takes a filename of URL of a blog feed and classifies the entries
8 def read(feed, classifier):
9
10     splitExpression = re.compile(r"<[>]+>")
11     num = 0
12     print "first 50 entry classification"
13     # Get feed entries and loop over them
14     f=feedparser.parse(feed)
15
16     for entry in f['entries'][0:50]:
17         num = num+1
18         print
19         print '_____'
20
21         # Print the contents of the entry
22         title = entry['title'].encode('utf-8').replace(" ", "")
23         print 'Title: '+title
24
25         summary = splitExpression.sub("", entry["summary"])
26         print summary
27
28         # Combine all the text to create one item for the classifier
29         fulltext = '%s\n%s' % (entry['title'], entry['summary'])
30         fulltext = fulltext.replace(" ", "")
31         predictedString = str(classifier.classify(fulltext))
32
33         # Print the best guess at the current category
34         print "predicted category", predictedString
35
36         # Ask the user to specify the correct category and train on that
37
38         actual = raw_input('Enter the actual category it belongs to: ')
39         classifier.train(fulltext, actual)
40
41     print "entering into another mode"
42     num = 50
43     print "other 50 entry classification"
44     # Get feed entries and loop over them
45     f=feedparser.parse(feed)
46     for entry in f['entries'][50:100]:
47         num = num+1
48         print
49         print '_____'
50
51         # Print the contents of the entry
52         title = entry['title'].encode('utf-8').replace(" ", "")
53         print 'Title: '+title
54
55         summary = splitExpression.sub("", entry["summary"])
```

```

56     print summary
57
58     # Combine all the text to create one item for the classifier
59     fulltext = '%s\n%s' % (entry['title'], entry['summary'])
60     fulltext = fulltext.replace(" ", "")
61     predicted= str(classifier.classify(fulltext))
62
63     # Print the best guess at the current category
64     print "predicted category",predicted
65
66     # Ask the user to specify the correct category and train on that
67
68     actualS =raw_input('Enter the actual category it belongs to: ')
69     feature = raw_input('Enter a string to classifier')
70
71
72     cprobabilty = round(classifier.cprob(feature ,predicted) ,4)
73     print cprobabilty
74
75     #classifier.train(fulltext ,actual)

```

2.2.3 entries.py

```
1 import docclass
2 import feedfilter
3
4 def main():
5     cl=docclass.fisherclassifier(docclass.getwords)
6     cl.setdb('bbokka.db')
7     print "testing the program"
8     feedfilter.read('test.xml',cl)
9
10 if __name__ == "__main__":
11     try:
12         main()
13     except KeyboardInterrupt:
14         sys.exit(1)
```

Title	Summary	Category
December 5	Greetings from here, where the three of us are still sick. Brandon told me that he counted it up in the bathtub this morning, and hes now been sick for 27 days.	personal
Like he did	The three of us have that hanger-onner of a virus thats going around. The past two nights, Ive coughed myself to sleep in the basement guest room, and as anyone whos ever coughed herself to sleep can tell you, its slow going.	personal
I got to go back	The first time I went to the Oklahoma Arts Institute at Quartz Mountain was in the summer of 1995, a few months after a fire destroyed the lodge, its rooms and dining hall and library.	place
Even on a good day	My mother has been in town since early this month. We dont often get this kind of extended time in the same place, and Id forgotten what a good cooking collaborator she is. She makes sure our wine glasses are never empty.	personal
Looking forward	I picked up a roll of film that I shot at Sam and Megans wedding last month, and maybe my friend at the lab did some wizardry with the negative scanner, but the whole roll has this glowy, ethereal light shining through it.	season
The satisfaction	In early September, a kind reader in north-central Idaho left me a comment. Her name was Michele, and her Italian prune plum tree was promising a bumper crop: did I want some? This kind of thing does not happen all the time, or ever, so I said (yelled) yes.	season
Every Tuesday	Whoa. I got sucked into a black hole for a bit there, a (very pleasant, very festive) black hole of weddings and out-of-town visitors. Somehow its now September 26, and Im glad to be alone tonight.	season
September 6	From the summer of 2006 until the early spring of 2011.	season
And: FUD-GESICLES	nnnI have a child who is about to be two years old. I have a lot of thoughts on the subject, but one thing I do not have a lot of thoughts about is a second birthday party.	personal
August 18	A couple of weekends ago, we packed up the better part of the restaurant kitchen, crammed it in the back of a pick-up, and drove two and a half hours east to cook an all-day anniversary party for a pair of longtime Delancey regulars. We rented a big house along the Wenatchee River.	place
July 22	A month of summer gone already! I dont want to think about it.	season
I promised	It hit 85 degrees in Seattle today.	season
July 6	We spent half of last week on Lopez Island, staying with friends at the home of friends-of-friends, breaking in our sun hats, making buildings out of driftwood.	personal
June 27	Friday! Its rainy here in Seattle.	season
Run with it	watching the worlds most subtle breeze blow through the branches of the neighbors tulip magnolia.	season
Right now	Hello from a train en route to Portland, Oregon!	place
June 1	Id been planning to put up a post tonight about some meatballs that June has been into lately (MEAT! MEAT! she yells;	recipe

Come on down	Hello from a plane somewhere between Minneapolis and San Francisco!	place
Our people	Im typing this post from my cousins kitchen table in Oakland, California, where June and I are visiting for a family baby shower and have stayed long enough to eat four slices of red velvet cake.	recipe
Maybe he's right	This is the fifth granola recipe Ive posted on this blog.	recipe
April 23	In the time since you and I last spoke, Ive been to New Jersey and back for my sister-in-laws wedding and, in the name of research, have eaten a lot of New Jersey bagels and New Jersey pizza.	recipe
That word is eat	June has mastered a new word, and that word is eat. Its one of many things I like about her.	personal
Book tour, housekeeping, and you will now have Hole songs stuck in your head	My publisher tells me that finished copies of Delancey, hot off the presses, are due to arrive in their offices early next week.	personal
Call it a meal	We have reached the point in winter, or spring, or whatever it is, when even I am tired of making, eating, and talking about soup.	recipe
A short leap	Our friend Ben is in town for a visit, and this past Friday, while we waited in line for lunch.	recipe
Felt like it	HELLOOOOOOOOO Im just off the plane from a week in Oklahoma City with June and my mother.	place
Always to acclaim	I typed most of this post on Valentines night, while Brandon worked at Delancey, slinging pizzas for all the lovers.	season
Book update, housekeeping, la la la	Thats the way I look at it in my better moments, the ones when Im not staring at the clock	personal
The right buttons	Before we get started: thank you for your comments on my previous post, and for sharing so many good tips and ideas about feeding families.	personal
In it together	First things first: if you dont want to read about kid stuff, you should skip this post. I wont mind. A few years ago, I totally would have skipped it.	personal
A good person to know	I first met Megan at a conference, I think? Im a real loser when it comes to conferences - crowds make me feel like hiding under furniture, and my brain is a wide-mesh sieve for faces and names.	personal
From now on	Our friend Ben was here last week. He arrived on Thursday, just in time for lunch.	personal
Their good work	Hello again! If I dont write a post tonight, I will have to do my real work, which is to read the final proofs of Delancey before it goes to print, and that is a terrifying prospect. So! La la laaaaaaaaaaaaaaa As it happens, combing through files of old film photos is also a great way to avoid work - and I can use the photos here! Behold: somewhat ancient photos that have nothing at all.	personal

Approximately a soup	First: RING THE BELLS! I HAVE A NEW CAMERA! Here at Wizenberg-Pettit World Headquarters, we are excited. And grabby. Second: we are also into soup, apparently, which is why Im going to tell you about yet another, our third soup in a row.	recipe
Please consider	So, how bored will you be if we talk about soup again? Ham Bone, Greens, and Bean Soup?	recipe.
But the soup	Would you look at that! While trying, and failing, to start this post about squash soup, I accidentally ate an entire chocolate chip cookie dough ball.	recipe
The days are twice as long	This time last week, I was in a wood stove-heated cottage with no Internet, no telephone, and no television, reading my sixth New Yorker of the day.	personal
It made an impression	After I brought it home, while I waited for my blood pressure to stabilize, I combed through the files that had been on my old computer and happened to find a document that I had forgotten, a recipe for a brown sugar clafoutis with pears. BROWN SUGAR CLAFOUTIS!	recipe
Over and over and over	It was a summer of greatest hits. Ive always been a redundant cook, content to repeat a couple dozen recipes over and over (and over and over and over).	recipe
Bottoms up	I planted about a half dozen plants back in late May	personal
A new reason	I just sat down, looked at the calendar, and noticed that its August 9th.	personal
Nine	I am typing this post from the back office at Delancey.	personal
A rare benefit	Does that ever happen to you? Do you do what I do and take a "break" to raid the walk-in at Delancey for chocolate chip cookie dough?	recipe
This thing is on	Maybe you will remember a day, more than two years ago now.	personal
June 28	Friday! Yessssssss. Its just after noon, and Im sitting in Essex, my office by day, with an imposing to-do list	personal
Told you so	Every so often, I encounter a recipe that makes me want to forgo the usual niceties of a post - the introduction, the story, the conclusion, the delicate foreplay - because that would only slow you down, when what you should really do is grab your shoes and make a list and run to the grocery store.	recipe
June 4	Last night, it occurred to me that I had inadvertently neglected to write down something important: that Junes head smells like strawberry jam. Ive thought about it for a long time, trying to make sure that was it, and now Im certain: not strawberries, but strawberry jam.	personal
Eureka	You women who manage to keep up smart, articulate blogs while raising young children? You women who manage to keep up smart, articulate blogs while working and raising young children and doing all that household stuff that most of us wind up doing? I throw myself at your feet. I dont have anything remotely original or insightful to say on the subject; I just think youre remarkable.	personal

I'm feeling darling	I have finally learned how to use the espresso machine that Brandon chased down on eBay and gave me for Christmas in 2011!	personal
April 7	Ive been feeling a little under the weather for the past few days, but I wanted to pop in. I promise not to breathe on you.	season

Table 1: Manually classified first 50 entries

Title	Summary	Actual	Predicted	cprob	(String	type(gram)
We have a rhythm	June is six months old. She has two teeth, monstrous thighs, and is my favorite person in the world.	personal	personal	0.0	june	1-gram
The usual	In the time since we last spoke, Ive revised the manuscript for my next book! Ive traveled alone with my five-month-old baby to a family wedding on the other side of the country!	personal	food	0.0	five-month-old	1-gram
A small revolution	You good, good people. Before I say another word, I want to thank you for your many comments, your e-mails, and the incredibly kind card - a real, three-dimensional paper card - that one of you sent to me at Delancey.	personal	personal	0.0	good people	2-gram
Ah ha	My father wasnt a writer, or not in the vocational sense, but he liked to play with words.	personal	personal	0.0	goofy	1-gram
As loud as I wanted	I have to get back to writing it down, or Ill forget.	personal	personal	0.0	esoteric vermouths	2-gram
I stand by my assessment	I know. The universe does not need another recipe for a banana baked something. And yet. Its a rainy Friday afternoon, and I have a couple of hours(!) home alone(!) to do whatever(!) I want(!), which is mostly to eat Hagen-Dazs coffee ice cream and stare at the wall	recipe	recipe	0.0	polite	1-gram
All right	I hope youre well. Things are all right here, Im relieved to report, especially given the fact that I had exactly zero experience with babies prior to birthing one three short weeks ago.	personal	personal	0.0	right	1-gram

We did it	I am very happy to announce that June is here.	season	season	0.0	we did	1-gram
Essex	I MADE IT! By which I mean, I managed to not go into labor before, in the middle of, or in the days immediately following the opening of Essex.	personal	recipe	1.0	immediately	1-gram
All of it	Hi, all. Ive received a number of concerned e-mails and comments in the past few days, wondering if the quiet around here meant that the baby has arrived, so let me say first: NO BABY YET.	personal	personal	0.0	finished writing	2-gram
Lately	Last night I went to the grocery store for milk and yogurt and the usuals, and the cashier smiled and commented on how much my belly has grown since she last rang me up.	personal	personal	0.0	grocery	1-gram
Let's wing it	Before I say anything else, I want to thank you for your kindness about my aunt. I was very nervous when I put up that post, but I felt much better for having written it.	personal	recipe	0.0	smooth transition	2-gram
She felt like cheering	I have three half-siblings.	personal	recipe	1.0	california	1-gram
You know me well	My manuscript is due on June 1.	personal	personal	0.0	meats	1-gram

No such thing	Earlier this week, I think it was, one of you kindly wrote to me, asking if I might do a post about what Ive been eating for lunch lately. The reader who wrote to me is pregnant, and there are a number of foods that us pregnant ladies are told to avoid, making quick, easy lunches hard to come by: no deli meats, no (uncooked) cured meats, no high-mercury fish (tuna, for example).	recipe	recipe	0.0	lunch lately	2-gram
The first night	Hi from here.I should clarify: not here exactly. These photos were taken at our dinner table, but Im currently sitting at my desk. I would much rather be there than here. Oh well.These shots are from the first night of 2012, with our friends Ben, Bonnie, and Sam. We had wanted to go crabbing and catch our New Years dinner ourselves.	personal	personal	0.0	brandon	1-gram
The best part of the job	I am supposed to be writing a manuscript, not baking rye crumble bars.	recipe	personal	0.0	crumble bars	2-gram
This is important	You people. YOU PEOPLE.Im still blinking in disbelief at the kindness youve shared with me and Brandon and Tiny Person Under My Shirt.	personal	personal	0.0	you people	2-gram
September 4	Hi.Ive never made an announcement like this before, and it feels very awkward and shouty to do it through a computer screen, so please be nice and pretend that were sitting in your living room.	personal	personal	0.0	sitting in your living	3-gram

Dear World,	I am writing to you, once again, from my friend Bens dining room. When I was here last August, writing my brains out, I had a hunch that a return visit might be helpful before my manuscript deadline.	personal	personal	0.0	manuscript	1-gram
I am here to tell you	First things first: thank you for the well wishes.	personal	personal	0.0	well	1-gram
You can count on me	I am writing to you today with a wool scarf wound around my entire upper body and a wool blanket tied at my waist.	personal	recipe	0.0	writing	1-gram
December 24	Whenever we spend the holidays with my relatives in northern California, we eat Dungeness crab on Christmas Eve.	season	personal	0.0	christmas	1-gram
In my better moments	About three weeks ago, I printed out all the drafts Ive written so far for my next book, and then I spent three weeks avoiding reading them.	personal	personal	0.0	three weeks	2-gram
To poach a pear	My mother is usually the one who makes poached pears.	personal	season	0.0	poached pears	2-gram
Its my specialty	Hi. I am writing this from my in-laws kitchen.	recipe	recipe	0.0	speciality	1-gram
October 31	Weve arrived at the end of my trip.	personal	personal	1.0	arrived	1-gram
October 29	To those of you who advised me to go to Scotland.	place	personal	0.0	advised	1-gram
October 19	My favorite church is in Paris, and its called Saint-Sulpice.	place	season	0.0	family	1-gram
October 22	I sat down to write this post last night at a kitchen table in Edinburgh.	personal	personal	0.0	last night	2-gram
October 16, 17	I love the mistakes that come with using film. Often, I like the mistakes more than the shots that turn out.	personal	season	0.0	mistakes	1-gram

October 15	Hi, friends.Im in Paris now.I know I just typed that as though it were nothing, but what I meant was: IM IN PARIS NOW.	place	personal	0.0	paris	1-gram
October 14	When I was eighteen, I took my first big trip without my parents, and before I left, my mother suggested that I pack a nice notebook to use as a journal.	personal	personal	0.0	eighteen	1-gram
You're lucky, she told me.	Hi, friends.This is not the post I had expected to write next. In my head, there were going to be cheers, an obscene number of exclamation points, and maybe a picture of the evening street outside our apartment in Paris.	place	personal	0.0	lucky	1-gram
Paris Diary	I first met my friend Maria in 2005.	place	personal	0.0	maria	1-gram
Out on this limb	I was planning to start this post with something about food, because thats the way these things usually work.	personal	personal	0.0	planning to start	3-gram
No frosting, no ceremony	"The citrus is coming! The citrus is coming!"	recipe	season	0.0	window open	2-gram
Not a tomato sandwich	So, I turned 33 a couple of days ago.	personal	personal	0.0	33 couple	2-gram
Ohio	I am writing to you today from my friend Bens dining room.	place	personal	0.0	ohio	1-gram
A real thrill	Im going to tell it to you straight. When I got to the last step of this recipe and looked at the tower of dirty bowls and saucepans in the sink, I thought, This had damn well better be the best frozen yogurt the universe has ever seen.	recipe	personal	0.0	quite	1-gram

It still surprises me	So. I think it would be fair to say that your comments on my last post made me very, very happy.	personal	personal	0.0	surprises me	2-gram
How we do what we do	Ive been out of town for the past week, helping with preparations for my cousins wedding in Oakland	place	recipe.	0.3459	town	1-gram
You don't forget	I may be too late for strawberries.	recipe	personal	0.0	strawberries	1-gram
I wish them a lifetime	I mentioned last week that I had been away at a wedding, Luisas wedding, and a number of you wrote that you were eager to see pictures.	personal	personal	0.0	lifetime	1-gram
I could talk about the weather	I woke up this morning and found the house entirely wrapped in fog.	personal	personal	0.0	woke up	2-gram
This one's coming with me	On a Sunday night in June, you are required, by cosmic law, to make strawberry shortcake.	recipe	personal	0.0	coming with	2-gram
Such is the power	Let the record show that, I, Molly Wizenberg, have, in this lifetime, made some ugly deviled eggs.	personal	personal	0.0	recordshow	2-gram
Your efforts will be rewarded	Listen, I know its a holiday weekend.	personal	personal	0.0	holiday weekend	2-gram
We have some progress	This city has taken its sweet time in getting rid of winter.	season	recipe	0.0	sweet	1-gram
pile it on	Let's get it out of the way right now: this egg salad, the one we are going to talk aboit today.	recipe	personal	1.0	pile	1-gram

Table 2: Entries classified based on fisher method

3 Question 3:

Assess the performance of your classifier in each of your categories by computing precision, recall, and F1. Note that the definitions of precisions and recall are slightly different in the context of classification; see:

http://en.wikipedia.org/wiki/Precision_and_recall#Definition_.28classification_context.29
and

http://en.wikipedia.org/wiki/F1_score

3.1 Approach

1. To calculate the precision, recall and F1 we need TP (True positive), TN (True negative), FP (False positive), FN (False negative).
2. Table 3 shows you the precision, recall and F1 values of each category.

Category	Precision	Recall	F1
personal	0.68	0.61	0.64
places	0	0	0
recipe	0.33	0.06	0.20
season	0.2	0.02	0.03

Table 3: Precision, Recall and F1 values for each category

References

- [1] Toby Segaran. *Programming Collective Intelligence*. O'Reilly Media, August 2007.
- [2] Molly Wizenberg. Blog. <http://orangette.blogspot.com/>.

□