170512_real_search_parsing

July 12, 2017

Compare lists of maybes and dips from 170705 and 170410 run

In [1]:
```
import pandas as pd
import numpy as np

first_path = '../results/real_search/170410_candidates_pass_1/human_labelle
first = pd.read_csv(first_path, names=['kicid','human_label'],
                    delimiter=', ', engine='python')
second_path = '../results/real_search/human_labelled_candidates.txt'
second = pd.read_csv(second_path, names=['kicid','human_label'],
                    delimiter=', ', engine='python')
```

In [2]:
```
for datestr, df in zip(['170410', '170705'], [first, second]):
    print('{:s} N labelled dip: {:d}'.format(
        datestr,len(df[df['human_label']=='dip'])))
    print('{:s} N labelled maybe_dip: {:d}'.format(
        datestr,len(df[df['human_label']=='maybe_dip'])))
    print('{:s} N labelled wtf: {:d}'.format(
        datestr,len(df[df['human_label']=='wtf'])))
```

```
170410 N labelled dip: 11
170410 N labelled maybe_dip: 39
170410 N labelled wtf: 0
170705 N labelled dip: 26
170705 N labelled maybe_dip: 84
170705 N labelled wtf: 16
```

In [3]:
```
first[first['human_label']=='dip']['kicid']
```

Out[3]:
```
1       6144827
8       8197406
10      7871200
12      5302006
24     11013201
32      7515679
43      9843451
99      9480977
```

```
273     6791604
320     5983351
372     7889628
Name: kicid, dtype: int64
```

```
In [4]: second[second['human_label']=='dip']['kicid']
```

```
Out[4]: 0       9790965
        1       6144827
        2      11013201
        5       8197406
        9       7871200
        43      5302006
        68     11303811
        116     9480977
        128    11811454
        139     8495415
        141     8330092
        221     9954225
        243     7941050
        277     5025261
        300    11135978
        327     5881838
        345    10934755
        458     7941635
        488     5642620
        531     9788113
        562     9788457
        603     8587078
        649     9705459
        713     9700181
        737     7889628
        782    11819135
        Name: kicid, dtype: int64
```

Which ones that were originally labelled dips are still?

```
In [5]: m = np.in1d(first[first['human_label']=='dip']['kicid'],
                   second[second['human_label']=='dip']['kicid'])
        first[first['human_label']=='dip']['kicid'][m]
```

```
Out[5]: 1       6144827
        8       8197406
        10      7871200
        12      5302006
        24     11013201
        99      9480977
        372     7889628
        Name: kicid, dtype: int64
```

```
In [6]: len(first[first['human_label']=='dip']['kicid'][m])

Out[6]: 7
```

And which are not?

```
In [7]: m = np.in1d(first[first['human_label']=='dip']['kicid'],
                    second[second['human_label']=='dip']['kicid'])
        first[first['human_label']=='dip']['kicid'][~m]

Out[7]: 32     7515679
        43     9843451
        273    6791604
        320    5983351
        Name: kicid, dtype: int64
```

7515679 is a harmonic of EB period

9843451 seems like noise from original dipsearchplot (bumps as big as dip – higher power harmonic peak is negative)

6791604 was identified b/c of the very big SNR per transit (not visible in the way I did dipsearchplot relative flux top panel...)

5983351 should be a maybe

```
In [8]: m = np.in1d(first[first['human_label']=='dip']['kicid'],
                    second[second['human_label']=='maybe_dip']['kicid'])
        first[first['human_label']=='dip']['kicid'][m]

Out[8]: 43    9843451
        Name: kicid, dtype: int64
```

But 5983351 was labelled in round 2 as noise. I agree that the LCs make it look like it, but the BLS spectrum usually wouldn't give such a high power peak. Probably should have been flagged as a maybe.

9843451 should really have been boosted from "maybes" to "dips" in round 2.

I think this just indicates the importance of bringing in another vetter...

1 New "dips":

```
In [9]: m = np.in1d(second[second['human_label']=='dip']['kicid'],
                    first[first['human_label']=='dip']['kicid'])
        second[second['human_label']=='dip']['kicid'][~m]

Out[9]: 0       9790965
        68     11303811
        128    11811454
        139     8495415
        141     8330092
        221     9954225
        243     7941050
```

```
277     5025261
300    11135978
327     5881838
345    10934755
458     7941635
488     5642620
531     9788113
562     9788457
603     8587078
649     9705459
713     9700181
782    11819135
Name: kicid, dtype: int64
```

In [ ]: