Assignment 3: Intro Quantitative Analysis

Jiaming Xu

Assignment of ECE1785H 2022W

1. Introduction & Motivation

As part of our final project, I have been doing interviews with many software engineers in big companies. To my surprise, many of them have contributed or are making contributions to the open-source projects. So I'm curious about what's the big companies attitude towards these open-source projects related experience.

Do we need to make more contributions to open-source projects if we want to apply for some companies which flavor open-source communities? Does age have an impact within? Does young people have to contribute more to open-source code to get their jobs? Will the employers ask us to contribute to open-source communities as our job if they encourage using/contributing to open-source projects? I think this problem matters because it can help me and many other software engineering students have a clearer view of this aspect before we try to seek for jobs. Hopefully, [1] [2] has mentioned part of these questions and data, so this report will explore these questions.

2. Research Questions

- RQ1: Do young people have to contribute more to open-source projects to get their position in companies which flavor open-source communities?
- RQ2: Do employees tend to contribute to open-source projects as their job in companies which encourage using to open-source projects?

3. Data Cleaning & Analysis

This is the dataset [3]İ used 'Python' to do the data cleaning and analysis. To do the data cleaning, I use the 'pandas' lib: the read_csv function to read the data and dropna function to drop the rows which has a NaN data.

For the first question, I choose the 'AGE' column to represent the age and working experience of the engineers, the 'EMPLOYER.POLICY.APPLICATIONS' column to show whether the companies flavor the open-source communities and the 'OSS.HIRING' column to explore if the experience helps them get their jobs.

For the second question, I choose the 'OSS.AS.JOB' column to represent whether they tend to contribute to open-source projects as their job, 'EMPLOYER.POLICY.APPLICATIONS' column to show their employers attitude towards open-source communities. In order for the data to fit into the linear regression model, I transformed the data into numbers. You can find the code in the appendix.

'OSS.HIRING': 'Very important':1,'Somewhat important':2,'Not at all important':3,'Not too important':4,'Not applicable-I hadnt made any contributions when I got this job.':5

'AGE': '17 or younger':1,'18 to 24 years':2,'25 to 34 years':3,'35 to 44 years':4,'45 to 54 years':5,'55 to 64 years':6,'65 years or older':7

'EMPLOYER.POLICY.APPLICATIONS':'Use of open source applications is encouraged.':1,'Use of open source applications is acceptable if it is the most appropriate tool.':2,'My employer doesnt have a clear policy on this.':3,'Not applicable':4,'Im not sure.':5,'Use of open source applications is rarely, if ever, permitted.':6

'OSS.AS.JOB'='Yes, directly- some or all of my work duties include contributing to open source projects.':1,'Yes, indirectly- I contribute to open source in carrying out my work duties, but I am not required or expected to do so.':2,'No.':3

4. Results

4.1. RQ1

The formula is $y = 0.04301915x_1 + 0.23456672x_2 + c$ where y is the predicted 'OSS.HIRING', x_1 is 'AGE', x_2 is 'EMPLOYER.POLICY.APPLICATIONS', c is a constant number. As is shown above, the less your companies show positive attitude to open-source communities, the older you are, your open-source experience is less likely to help you get the job. Controlling the age x_1 to be the same, we can see that y is increasing with x_2. It means the companies which do not like open-source projects pay less attention to the related working experience when hiring engineers. Controlling the policy x_2 to be the same, we can see that y is increasing with x_1. It means the older you are, the open-source projects will help you less to get your job.

4.2. RQ2

The formula is $y = 0.20563629x + c$ where y is the predicted 'OSS.AS.JOB', x is 'EMPLOYER.POLICY.APPLICATIONS',c is a constant number. [[0.20563629]]

The result shows that if your company has a positive attitude to open-source projects, you are more likely to contribute to open-source project as your paid job or part of it.

5. Conclusion

In conclusion, if we want to enter a company, it helps a lot to know whether this company flavors the open-source projects. If it does, we should contribute more to the open-source projects which can help us greatly in the application. And as young students, we should do

x1	x2	y
1	1	2.293055827077731
1	2	2.5276225515024704
1	3	2.76218927592721
1	4	2.9967560003519496
1	5	3.231322724776689
1	6	3.4658894492014287
2	1	2.3360749754228634
2	2	2.5706416998476027
2	3	2.805208424272342
2	4	3.039775148697082
2	5	3.2743418731218217
2	6	3.5089085975465606
3	1	2.3790941237679952
3	2	2.613660848192735
3	3	2.8482275726174744
3	4	3.0827942970422137
3	5	3.3173610214669536
3	6	3.551927745891693
4	1	2.4221132721131275
4	2	2.656679996537867
4	3	2.8912467209626067
4	4	3.125813445387346
4	5	3.3603801698120854
4	6	3.5949468942368252
5	1	2.46513242045826
5	2	2.6996991448829992
5	3	2.9342658693077386
5	4	3.1688325937324784
5	5	3.4033993181572177
5	6	3.637966042581957
6	1	2.5081515688033917
6	2	2.7427182932281315
6	3	2.977285017652871
6	4	3.2118517420776103
6	5	3.44641846650235
6	6	3.6809851909270894
7	1	2.551170717148524
7	2	2.7857374415732634
7	3	3.020304165998003
7	4	3.2548708904227426
7	5	3.489437614847482
7	6	3.7240043392722217

x	y
1	1.9315087682434604
2	2.1371450584278575
3	2.342781348612254
4	2.5484176387966513
5	2.7540539289810484
6	2.959690219165445

more open-source contributions for our future jobs. What's more, in companies which have a positive attitude towards open-source projects, we are more likely to contribute to open-source projects as our job.

参考文献

[1] F. Zlotnick, "Github open source survey 2017," http://opensourcesurvey.org/2017/, Jun. 2017.

[2] R. S. Geiger, "Summary analysis of the 2017 github open source survey," *CoRR*, vol. abs/1706.02777, 2017. [Online]. Available: http://arxiv.org/abs/1706.02777

[3] F. Zlotnick, "Github open source survey 2017 [data set]," https://doi.org/10.5281/zenodo.806811, Jun. 2017.