

Paper Outline

Team 9: G.Zhao, J.Bernstein, Y.Li

May 6, 2022

1 Abstract

2 Introduction

2.1 Topic

2.2 Dataset

3 EDA

We got the 20,122 comments data from Bitcoin community from GitHub. After getting this data, we used the package to get the leadership corresponding to each comment, which corresponds to LD1-LD6 respectively.huang2022identifying If there is no leadership corresponding to this comment, it will be marked as N. In this project, we consider problems from two perspectives, one is all issues. The other is all the developers.[?]

3.1 What is the data

As the measurements of the contribution of the community, leadership status is not necessarily based on an OSS community position or designated authority. we adopt the Yukl’s definition of leadership[?] as “influence exerted over other people to guide, structure, and facilitate relationships in a group. Utilize the NLP algorithm and packed into iLead, labeled 7 categories of the leadership (LD1/LD2...LD6/N)[?, ?] behaviors based on these existing leadership studies, we develop an automatic algorithm to merge and consolidate different pattern sets extracted from multiple projects into a final pattern ranking list. When applying iLead for leadership identification, the final pattern ranking list can be used to automatically match new issue comments and identify corresponding leadership behaviors. The evaluation is conducted on 5 OSS project in GitHub topic that varies from atom/bitcoin/brew/ember/sklearn, and more than 100000 issue comments. Results show that iLead can achieve a median precision of 0.82 and recall of 0.78,outperforming ten machine/deep learning baselines.[?]

- LD1 (Proposal). This category refers to the idea proposal type of leadership behaviors in issue discussion, which maps to the intellectual stimulation/inspiration category in existing leadership studies
- LD2 (Redirection). This category refers to issue comments redirecting attentions to a new topic (e.g., concepts, processes, or resources) or to a more relevant information sources, either internal (e.g., a related issue within the OSS project) or external (a site outside the OSS project)

- LD3 (Confirmation). This category refers to the leadership function in supporting and coordinating an issue or opinion expression based on voting or shared perception among developers.
- LD4 (Inquiry). This category refers to cases where developers ask for more clarification question about the issue being reported/discussed, e.g., the project version and reproduce steps, or provide confirmation to former opinions, to facilitate the issue moving forward.
- LD5 (Operation) and LD6 (Volunteer) represent two distinct types of administrative functions during issue life cycle. LD5 refers to comments related to either closing or reopening an issue, and LD6 relates to volunteering in undertaking some wrap-up bug fixing actions, respectively.

3.2 Cleaning

3.3 Preparation

4 Modeling

4.1 Models used

4.2 Prediction

4.2.1 Is someone a leader? (binary classification?)

1. Based on a few features like number of comments, time between issue opening and first comment, time between last comment and closing of issue, number of comments on an issue, etc
- 2.
- 3.