An empirical study of emoji use in software development communication

Shiyue Ronga, Weisheng Wanga, Umme Ayda Mannan, Eduardo Santana de Almeida, Shurui Zhoud, Iftekhar Ahmed

OBJECTIVE

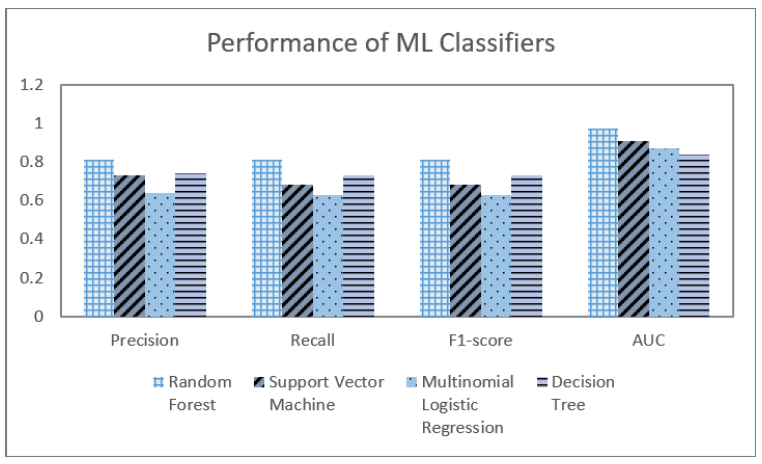
Understanding emoji usage in software development communication

- **RQ1** How well can we identify the intentions of emoji usage using machine learning?
- **RQ2** What intentions do developers have while using emojis during conversations?
- **RQ3** When do developers use emojis the most during the conversation?
- **RQ4** Do core and non-core developers use emojis differently?

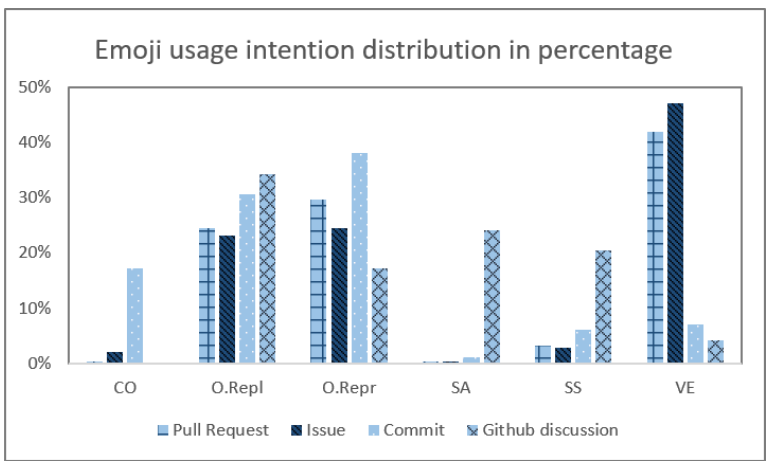
METHOD

We present a large-scale empirical study on the intention of emoji usage conducted on 2,712 Open Source Software (OSS) projects. We build a machine learning model to automate classifying the intentions behind emoji usage in 39,980 posts. We also surveyed 60 open-source software developers from 17 countries to understand developers' perceptions of why and when emojis are used.

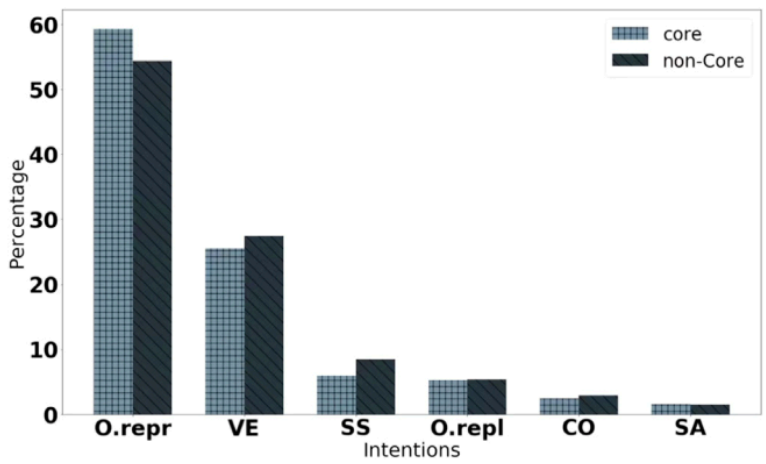
RESULTS



Random Forest Classifier can identify the intention of the emojis with an average AUC of 0.97.



Request Forerst Classifier used on the dataset consisting of four types of data 30,046 pull request comments, 5,908 issue comments, 3,778 commit messages and 248 discussion comments, to classify them in 6 Intention Categories : Content Organization (CO), Object Replacement (O. Repl), Object Representation (O. Repr), Sentimental Addition (SA), Sentimental Strengthening (SS) and Visual Enchancement (VE).



Open source contribution follows a power law, where 20% of contributors are responsible for 80% of the contributions. Following this rule, a developer is considered as **core** if the they are among the top 20% of developers in that project, based on the number of commits authored. Otherwise, the developer is **non-core**. Findings there are 91,094 core developers (22.2%) and 318,817 non-core developers (77.8%).

CONCLUSIONS

Observation 1

Random Forest Classifier can identify the intention of the emojis with an average AUC of 0.97.

Observation 2

Developers' intention of using emojis differs based on which channel emojis are used on, while pull request and issue shares the same distribution due to their similar functionalities on GitHub.

Observation 3

Emojis tend to be used during every phase in a conversation on GitHub.

Observation 4

Developers' role as a core or non-core contributor of a project does not impact the number of emoji usages and the intention of emoji usage.

MSc Applied Informatics
SD104-Software Engineering Technologies
Anastasios Chrysochoidis

HELLENIC REPUBLIC
UNIVERSITY OF MACEDONIA