

# Exploration and Exploitation of Developers' Sentimental Variations in Software Engineering

Md Rakibul Islam, University of New Orleans, New Orleans, LA, USA

Minhaz F. Zibran, University of New Orleans, New Orleans, LA, USA

## ABSTRACT

Software development is highly dependent on human efforts and collaborations, which are immensely affected by emotions. This paper presents a quantitative empirical study of the emotional variations in different types of development activities (e.g., bug-fixing tasks), development periods (i.e., days and times) and in projects of different sizes involving teams of variant sizes. The study also includes an in-depth investigation of emotions' impacts on software artifacts (i.e., commit messages) and exploration of scopes for exploiting emotional variations in software engineering activities. This work is based on careful analyses of emotions in more than 490 thousand commit comments across 50 open-source projects. The findings from this work add to our understanding of the role of emotions in software development, and expose scopes for exploitation of emotional awareness in improved task assignments and collaborations.

## KEYWORDS

Analysis, Development, Emotion, Maintenance, Sentiment, Software

## INTRODUCTION

Emotions are inseparable part of human nature, which influence people's activities and interactions, and thus emotions affect task quality, productivity, creativity, group rapport and job satisfaction (Choudhury & Counts, 2013; Feldt, Angelis, Torkara, & Samuelsson, 2010; Palacios, López, Crespo, & Acosta, 2010). Software development, being highly dependent on human efforts and interactions, is more susceptible to emotions of the practitioners. Hence, a good understanding of the developers' emotions and their influencing factors can be exploited for effective collaborations, task assignments (Dewan, 2015), and in devising measures to boost up job satisfaction, which, in turn, can result in increased productivity and projects' success (Denning, 2012).

Several studies have been performed in the past for understanding the role of human aspects on software development and engineering. Some of those earlier studies address when and why employees get affected by emotions (Choudhury & Counts, 2013; Guzman, Azócar, & Li, 2014; Guzman & Bruegge, 2013; Pletea, Vasilescu, & Serebrenik, 2014; Tourani, Jiang, & Adams, 2014), whereas some other work address how (Graziotin, Wang, & Abrahamsson, 2013; Khan, Brinkman, & Hierons, 2010; Lesiuk, 2005; Murgia, Tourani, Adams, & Ortu, 2014; Wrobel, 2013) the emotions impact the employees' performance at work. Despite those earlier attempts, software engineering practices still lack theories and methodologies for addressing human factors such as, emotions, moods and feelings (Graziotin, Wang, & Abrahamsson, 2015; Guzman & Bruegge, 2013). Hence,

DOI: 10.4018/IJSI.2016100103

Copyright © 2016, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

the community calls for research on the role of emotions in software engineering (Khan et al., 2010; Palacios et al., 2010; Shaw, 2004).

Some software companies try to capture the developers' emotional attachments to their jobs by means of traditional approaches such as interviews and surveys (Wrobel, 2013). Capturing emotions with the traditional approaches is more challenging for projects relying on geographically distributed team settings and voluntary contributions (e.g., open-source projects) (Destefanis, Ortu, Counsell, Marchesi, & Tonelli, 2015; Guzman et al., 2014). Thus, to supplement or complement those traditional sources, software artifacts such as the developers' commit comments/messages have been identified for the extraction of important information including developers' emotional states (Guzman et al., 2014; Guzman & Bruegge, 2013; Pletea et al., 2014).

This work conducts a study of the polarity (i.e., positivity, negativity, and neutrality) of emotions expressed in commit messages as posted by developers contributing to open-source projects. In particular, following five research questions are addressed.

**RQ1:** Do developers express different levels (e.g., high, low) and polarity (i.e., positivity, negativity, and neutrality) of emotions when they commit different types (e.g., bug-fixing, new feature implementation, refactoring, and dealing with energy and security-related concerns) of development tasks?

- If development tasks can be distinguished at which the developers' express high negative emotions, low positive emotions, or an overall low emotional involvement, stipulating measures can be introduced to emotionally influence the emotions of the developers working on those particular types of development tasks resulting in higher success rate.

**RQ2:** Can a group of developers be distinguished who express more emotions (positive or negative) in committing a particular type (e.g., bug-fixing) of tasks?

- Programmers who develop in them positive emotions while carrying out a given development task can be more efficient and quicker in completing the task (Murgia et al., 2014) resulting in reduced software cost. Thus, distinguishing a group of practitioners having positive emotional attachment to a particular task can be useful in effective task assignments.

**RQ3:** Do the developers' polarity (i.e., positivity, negativity, and neutrality) of emotions vary in different days of a week and in different times of a day?

- If particular days and times can be identified when developers' express significant negative emotions, then managers can take motivating steps to boost up the developers' positive feelings on those days and times. Guzman et al. (2014) reported that commit comments posted on Mondays tend to have more negative emotions. This work wants to verify their claim using a substantially larger data-set.

**RQ4:** Do the developers' emotions have any impact on the lengths of commit comments they write?

- Commit messages are pragmatic means of communication among the developers contributing to the same project. Ideally, commit comments contain important information about the underlying development tasks, and the length of developers' work description is an indication of the description quality (Maalej & Happel, 2009). If any relationship can be found between the developers' emotional state and the lengths of commit comments, then project managers can take steps to stimulate the developers' emotional states to get high quality commit comments containing enough contextual information.

**RQ5:** Do the sizes of the projects and development teams have any impacts on the developers' emotional states at work?

- Typically, large software projects require many developers and long development periods to complete. The challenge of working with a large code-base in collaboration with a large number of developers can cause negative emotions in the developers. In addition, large projects are typically more complex for staffing, assigning jobs, and sometimes those projects are less visible to anticipate critical risks beforehand. All these factors contribute

to make project management difficult and the managerial decisions also largely affect the emotions of the developers. A better understanding of the impacts of the sizes of projects and development teams can be exploited to manage developers' emotions effectively resulting in higher productivity and improved job satisfaction.

This paper is a significant extension to a recently published empirical study (Islam & Zibran, 2016). This work includes a deeper analysis of the research questions and a broader discussion of the existing literature relating the contrasts and contributions of this particular study. An additional research question (RQ5) and corresponding analyses are also included for better completeness of the work.

## METHODOLOGY

To address the aforementioned research questions, in this work, emotions are extracted from the developers' commit messages using *SentiStrength* (Thelwall, Buckley, & Paltoglou, 2012), which is a state-of-the-art sentiment analysis tool. *SentiStrength* was previously used for similar purposes (Garcia, Zanetti, & Schweitzer, 2013; Guzman & Bruegge, 2013; Tourani et al., 2014) and was reported to be good candidate for analyzing emotions in commit comments (Guzman, AzÓcar, & Li, 2014). The following subsections start with a brief introduction of sentiment analysis with *SentiStrength* and then, present the definitions of the metrics used in this work. The remaining last two subsections discuss about tuning of *SentiStrength* for software engineering context and data collection approaches used in this study. The procedural steps of this empirical study are summarized in Figure 1.

### Sentiment Analysis

Sentiment analysis using *SentiStrength* on a given piece of text (e.g., a commit message)  $c$  computes a pair  $\langle \rho_c, \eta_c \rangle$  of integers, where  $+1 \leq \rho_c \leq 5$  and  $-5 \leq \eta_c \leq -1$ . Here,  $\rho_c$  and  $\eta_c$  respectively represent the positive and negative emotional scores for the given text  $c$ .

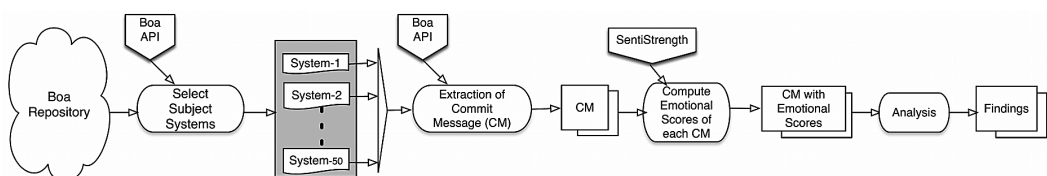
A given text  $c$  is considered to have positive emotions if  $\rho_c > +1$ . Similarly, a text is held containing negative emotions when  $\eta_c < -1$ . Note that, a given text can exhibit both positive and negative emotions at the same time, and a text is considered emotionally neutral when the emotional scores for the text appear to be  $\langle 1, -1 \rangle$ . Further details about the sentiment analysis algorithm of *SentiStrength* and the interpretation of its outputs can be found elsewhere (Thelwall et al., 2012).

### Metrics

To carry out analyses for deriving the answers to the research questions, following metrics are formulated. Given a set of commit messages, two subsets  $C_+$  and  $C_-$  can be obtained, which are defined as follows:

$$C_+ = \{c | c \in C, \rho_c > +1\} \text{ and } C_- = \{c | c \in C, \eta_c < -1\}.$$

Figure 1. Procedural steps of the empirical study



**Mean Positive Emotional Score** for a set  $C$  of commit messages, denoted as  $P(C)$ , is defined as:

$$P(C) = \frac{\sum_{c \in C_+} \rho_c}{|C_+|} \quad (1)$$

**Mean Negative Emotional Score** for a set  $C$  of commit comments, denoted as  $N(C)$ , is defined as follows:

$$N(C) = \frac{\sum_{c \in C_-} \eta_c}{|C_-|} \quad (2)$$

**Cumulative Emotional Score** for a particular commit message  $c$ , denoted as  $T(c)$ , is defined as follows:

$$T(c) = \rho'_c + \eta'_c \quad (3)$$

where,

$$\rho'_c = \begin{cases} \rho_c, & \text{if } \rho_c > +1 \\ 0, & \text{otherwise.} \end{cases}$$

$$\eta'_c = \begin{cases} |\eta_c|, & \text{if } \eta_c < -1 \\ 0, & \text{otherwise.} \end{cases}$$

### Tuning of *SentiStrength*

The sentiment analysis tool *SentiStrength* was reported to have 60.7% precision for positive texts and 64.3% for negative texts (Thelwall et al., 2012). To the best of the authors' knowledge, all such sentiment analysis tools including *SentiStrength* are highly dependent on the polarities of individual words in a given text in computation of its emotional scores. *SentiStrength* was originally trained on documents on the social web. In a technical field such as software engineering, commit messages include many keywords, which have polarities in terms of dictionary meanings, but do not really express any emotions in their technical context. For example, 'Super', 'Support', 'Value' and 'Resolve' are English words with known positive emotions, while 'Dead', 'Block', 'Default', and 'Garbage' are known to have negative emotions, but neither of these words really bear any emotions in software development artifacts. Those are simply some domain specific technical terms with especial contextual meanings.

To save *SentiStrength's* computation of emotional scores from being misled by such technical terms, the tool is tuned for application in the software engineering context. Based on manual investigation, experience, and literature review (Pletea et al., 2014; Tourani et al., 2014), a total of 49 terms are identified, which can be misinterpreted by *SentiStrength*. These misleading terms are: 'Arbitrary', 'Block', 'Bug', 'Conflict', 'Constraint', 'Corrupt', 'Critical', 'Dynamic', 'Dead', 'Death', 'Default', 'Defect', 'Defensive', 'Disabled', 'Eliminate', 'Error', 'Exceptions', 'Execute', 'Failure', 'Fatal', 'Fault', 'Force', 'Garbage', 'Greater', 'Inconsistency', 'Interrupt', 'Kill', 'Like', 'Obsolete',

'Pretty', 'Redundant', 'Refresh', 'Regress', 'Resolve', 'Restrict', 'Revert', 'Safe', 'Security', 'Static', 'Super', 'Support', 'Success', 'Temporary', 'Undo', 'Value', 'Violation', 'Void', 'Vulnerable' and 'Wrong'. *SentiStrength* provides the flexibility to modify its existing lexicons' emotional interpretation to customize it for a target context (i.e., software engineering, in this work). For this empirical study, those technical jargons are neutralized for *SentiStrength's* interpretation, as such was also suggested in earlier studies in the area (Pletea et al., 2014; Tourani et al., 2014).

Having *SentiStrength* tuned according to the procedure described above, the impact of the tuning is manually verified using a random sample of 200 commit messages extracted from Boa (Dyer, Nguyen, Rajan, & Nguyen, 2013). The manual verification found a 26% increase of precision as determined by comparing *SentiStrength's* computation of emotional polarities with subjective human interpretation over each of the 200 commit messages. Thus, this work uses an improved instance of *SentiStrength*, which is tuned for use in software engineering context.

## Data Collection

The commit messages for open-source projects, as required for this study, are obtained through Boa (Dyer et al., 2013). Boa is a recently introduced infrastructure with a domain specific language and public APIs to facilitate mining software repositories. The largest data-set is used (as of June 2016) from Boa, which is categorized as "full (100%)" and consists of more than 7.8 million projects collected from GitHub before September 2015.

From this large data-set, top 50 projects are selected, which have the highest number of commits. This study includes all the commit messages in these projects, which constitute 490,659 commit comments. Associated information such as, committers, commit timestamps, types of underlying work, revisions and project IDs are kept in a local database for convenient access and query. For each of the commit messages, emotional scores are computed using the tuned *SentiStrength* tool. Table 1 shows some examples of emotional and neutral commit comments in the dataset and computation of their emotional scores.

## ANALYSIS AND FINDINGS

After collecting required data, comprehensive analyses are conducted to answer the research questions RQ1, RQ2, RQ3, RQ4 and RQ5. The analyses and derivation of the answers to these research questions are presented in the following subsections.

### Emotional Variations in Different Task Types

Developers' emotions are investigated to determine whether those vary based on their involvements in five different types of software development tasks: (a) bug-fixing tasks, (b) new feature implementation, (c) refactoring, (d) energy-aware development and (e) security-related. Here first three types of tasks mentioned above are self-explanatory. The fourth one (i.e., energy-aware development) deals with software issues with consumption of energy, measured in terms of usage of resources such as processing power and memory. Energy-aware development is a recent important topic in the area of green computing research. The fifth category, software security, has also drawn tremendous importance in the community (Pletea et al., 2014) for various reasons. Categorization of development tasks in this manner is also found in earlier studies (Ayalew & Mguniin, 2013; Chowdhury & Hindle, 2016; Pinto, Castor, & Liu, 2014) in software engineering research.

#### Task-based Characterization of Commits

To distinguish commits dealing with bug-fixing tasks, this work invokes Boa's public API, which readily indicates whether a commit message is associated with bug-fixing task, or not.

To characterize rest of the task types, keywords-based searching mechanism is used. For example, to identify energy-aware commit messages, a list of related keywords is selected to characterize those

Table 1. Examples of commit comments and computation of their emotional scores

| Boa Project ID | Commit/Revision ID: Commit Comment ( $c$ )  | Emotional Scores |          |        |
|----------------|---|------------------|----------|--------|
|                |   | $\rho_c$         | $\eta_c$ | $T(c)$ |
| 7770259        | cc6e6d82661f59dc9d9bbb6b7d7c75b748a9a590: Fixes #1721 Another excellent patch from Daniel Siwiec, OPENEJB-1623: Example: Sharing of @ ApplicationScoped beans Thank you, Daniel!  | 5                | -1       | 5      |
| 6888786        | 1a633d7ee463482c9fc75d67a3a12d6ed21a1d86: add really horrible hacks to integrate with osx this code is not just ugly, it hurts my eyes  | 1                | -5       | 5      |
| 11814891       | 01845191185d0a14960f1542ac77f512f8749514 : a bit more detailed test; hope this avoids some reflection searches in FF emulation and makes the monster faster in special situations | 3                | -2       | 5      |
| 7709206        | 000b50baa73b3f25ae107886cd53aeda591b9f05:<br>table: log match and tuning parameters via match progress indicator  | 1                | -1       | 0      |

in commit messages. A commit message is considered as energy-aware commit, if the commit message contains any of the selected keywords. Table 2 lists the set of keywords used to characterize commit messages associated with the energy-aware, new-feature-development, refactoring, and security-related tasks. The character ‘\*’ in each keyword works as a wildcard, i.e., a query will select those commits messages, which contain at least one of these keywords, regardless of the beginning or the end of the commit message. Indeed, a developer may perform refactoring while fixing a bug. Thus, a commit message can be characterized relevant to more than one category of tasks.

Table 2. Lists of keywords used for task-based characterization of commit messages

|  |
|--|
| <b>Energy-aware development</b>  |
| <b>Keywords:</b> *energy consum*, *energy efficien*, *energy sav*, *save energy*, *power consum*, *power ecien*, *power sa*, *save power*, *energy drain*, *energy leak*, *tail energy*, *power efficien*, *high CPU*, *power aware*, *drain*, *no sleep*, *battery life* and *battery consum*.  |
| <b>New feature development</b>   |
| <b>Keywords:</b> *add* and *new feature*   |
| <b>Refactoring</b>   |
| <b>Keywords:</b> *refactor* and *code clean*   |
| <b>Security-related development</b>  |
| <b>Keywords :</b> *accesspolicy*, *accessrole*, *access-policy*, *access-role*, *accesspolicy*, *accessrole*, *aes*, *audit*, *authentic*, *authority*, *authoriz*, *biometric*, *blacklist*, *black-list*, *blacklist*, *blacklist*, *cbc*, *certificate*, *checksum*, *cipher*, *clearance*, *confidentiality*, *cookie*, *crc*, *credential*, *crypt*, *csrf*, *decode*, *defensiveprogramming*, *defensive-programming*, *delegation*, *denialofservice*, *denial-of-service*, *diehellman*, *dmz*, *dotfuscator*, *dsa*, *ecdsa*, *encode*, *escrow*, *exploit*, *firewall*, *forge*, *forgery*, *gssapi*, *gss-api*, *gssapi*, *hack*, *hash*, *hmac*, *honeypot*, *honey-pot*, *honeypot*, *inject*, *integrity*, *kerberos*, *ldap*, *login*, *malware*, *md5*, *nonce*, *nss*, *oauth*, *obfuscate*, *openauth*, *open-auth*, *openauth*, *openid*, *owasp*, *password*, *pbkdf2*, *pgp*, *phishing*, *pki*, *privacy*, *privatekey*, *private-key*, *privatekey*, *privi-lege*, *publickey*, *public-key*, *publickey*, *rbac*, *rc4*, *repudiation*, *rfc2898*, *rfc-2898*, *rfc2898*, *rijndael*, *rootkit*, *rsa*, *salt*, *saml*, *sanitiz*, *secur*, *sha*, *shell code*, *shell-code*, *shellcode*, *shibboleth*, *signature*, *signed*, *signing*, *singsign-on*, *singlesignon*, *single-sign-on*, *smartassembly*, *smart-assembly*, *smartassembly*, *snif*, *spam*, *spnego*, *spoofing*, *spyware*, *ssl*, *sso*, *steganography*, *tampering*, *trojan*, *trust*, *violate*, *virus*, *whitelist*, *white-list*, *whitelist*, *x509*, *xss* |

Note that, to identify energy-aware commits, the same set of keywords was previously used in the literature (Chowdhury & Hindle, 2016; Malik, Zhao, & Godfrey, 2015; Moura, Pinto, Ebert, & Castor, 2015; Pinto et al., 2014). The keywords for characterization of commits for new feature development and code refactoring, as listed in Table 2, were also used for the same purpose by Ayalew and Mguniini (2013). Similarly, for identifying commits pertaining to software security-related development, Pletea et al. (2014) also used the same set of keywords as listed in Table 2.

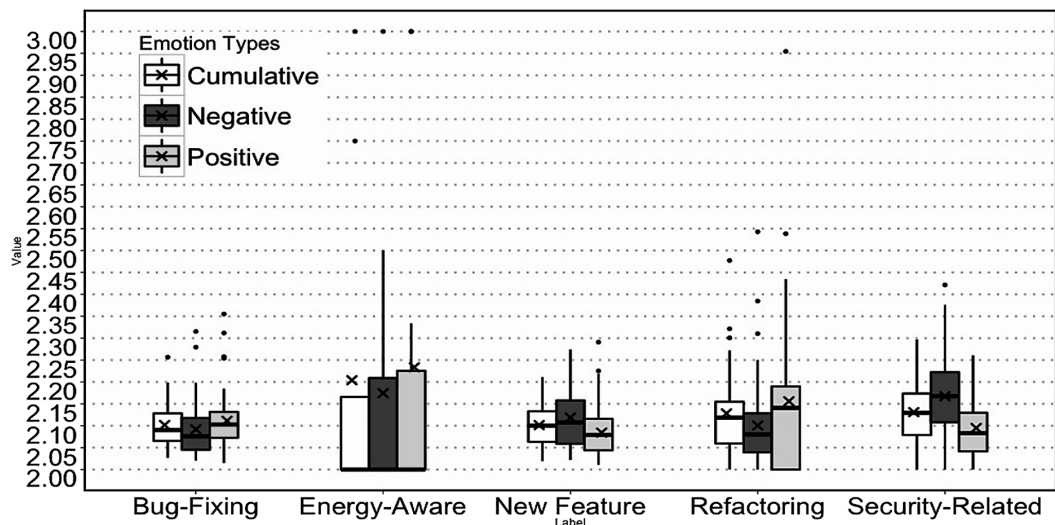
### Investigation

The numbers of commit messages found relevant to each of the five categories of development tasks are presented in the second column from left in Table 3. The boxplot in Figure 2 presents the distribution of mean positive, negative, and cumulative emotional scores in each type of task for each of the 50 projects. An 'x' mark in a box in the boxplot indicates the mean emotional scores over all the projects.

**Table 3. Commits for different categories of tasks and the results of MWW tests (P-values) between positive and negative emotions in them**

| Task Categories  | Number of Commits | <i>P</i> -values | Statistically Significant? |
|------------------|-------------------|------------------|----------------------------|
| Bug-Fixing       | 117,249           | 0.03288          | Yes ( $P < \alpha$ )       |
| Energy-Aware     | 182               | 0.39743          | No ( $P > \alpha$ )        |
| New Feature      | 89,019            | 0.00256          | Yes ( $P < \alpha$ )       |
| Refactoring      | 5,431             | 0.04006          | Yes ( $P < \alpha$ )       |
| Security-Related | 33,409            | 0.0              | Yes ( $P < \alpha$ )       |

**Figure 2. Distribution of mean positive, negative, and cumulative emotional scores in commits messages dealing with different types of tasks**



As observed in Figure 2, emotional scores (positive, negative and cumulative) for energy-aware commit messages are much higher than those in commit messages for four other tasks. Although the cumulative emotional scores in security-related commit messages are higher compared to bug-fixing and new feature implementation tasks, there is not much variations among the commits for rest other tasks. To verify the statistical significance of these observations, Mann-Whitney-Wilcoxon (*MWW*) (Ramsey & Schafer, 2002) tests are conducted (with  $\alpha = 0.05$ ) between the distributions of mean cumulative emotional scores in commit messages for each possible pair of development tasks. The results of the *MWW* tests (*P*-values) are presented in Table 4. The *P*-values reported by the tests, as compared with  $\alpha$ , suggest statistical significance of the observations.

Again, it can be seen in Figure 2 that the commit messages, which are posted during the implementation of new features and security-related tasks, show more negative emotions than positive ones. Opposite observations are evident for commit messages for three other types of tasks. To verify the statistical significance of these observations in the variations of polarity (positivity and negativity) of emotions, for each of the five types of development tasks, *MWW* tests are separately conducted between the mean positive and negative emotional scores of commit messages. The results of the *MWW* tests are presented in the right-most two columns in Table 3. The *P*-values of tests, as compared with  $\alpha$ , suggest statistical significance of these observations for bug-fixing, new feature implementation, refactoring and security-related tasks, but not for the energy-aware development tasks. These indicate that new feature development and security-related development tasks are not perceived (by the developers) as rewarding as the three other tasks.

Based on observations and statistical tests, the answer to the research question RQ1 is derived as in Box 1.

### Emotional Variations in Bug-Fixing Tasks

It is natural that different developers have different expertise, comfort-zones, and interests with respect to types of tasks. The research question RQ2 addresses the possibility of distinguishing a set of developers who particularly express positive emotions at the particular type of task at hand. In addressing the research question RQ2, this work chooses bug-fixing tasks as a representative to any particular type of tasks and continues as such.

Across all the projects, 20 developers are distinguished, who are the authors of the bug-fixing commit messages having the highest positive mean emotional scores. Let  $D_p$  denote the set of these

Table 4. Results of *MWW* tests (*P*-values) between cumulative emotional scores of commit messages dealing with different types of tasks

| Task Categories  | Bug-Fixing     | Energy-Aware | New Feature    | Refactoring | Security-Related |
|------------------|----------------|--------------|----------------|-------------|------------------|
| Bug-Fixing       | -              | 0            | 0.89656        | 0.75656     | <b>0.01314</b>   |
| Energy-Aware     | 0              | -            | 0              | 0           | 0                |
| New Feature      | 0.89656        | 0            | -              | 0.71884     | <b>0.02202</b>   |
| Refactoring      | 0.75656        | 0            | 0.71884        | -           | 0.29834          |
| Security-Related | <b>0.01314</b> | 0            | <b>0.02202</b> | 0.29834     | -                |

#### Box 1

**Ans. to RQ1:** Developers express significantly high positive and negative emotions almost equally in committing energy-aware tasks. For bug-fixing and refactoring tasks, positive emotions are significantly higher than negative emotions. On the other hand, significant higher negative emotions can be observed in developers during implementation of new feature and security-related tasks.



20 developers. Similarly, another set  $D_n$  consisting of 20 developers are formed, who are the authors of bug-fixing commit comments having the highest negative mean emotional scores. By the union of these two sets, the researchers obtain a set  $D$  of 30 developers who are authors of bug-fixing commits with the highest mean positive or negative emotional scores. Mathematically,  $D = D_p \cup D_n$ .

These 30 developers are the authors of 112,462 commits messages among which 32,088 are bug-fixing commits. For each of these 30 developers, a ratio  $R(d)$  is computed as follows:

$$R(d) = \frac{P(C_d)}{N(C_d)} \text{ where, } d \in D \quad (4)$$

Here,  $C_d$  denotes the set of bug-fixing commit comments posted by developer  $d$ . Notice that, for a particular developer  $d$ , the ratio  $R(d)$  close to 1.0 indicates that the positive and negative emotions are almost equal for the developer  $d$ . If  $R(d)$  is much higher than 1.0, the developer  $d$  shows more positive emotions at bug-fixing tasks compared to negative emotions. The opposite holds when  $R(d)$  is much lower than 1.0. However, a threshold scheme seems necessary to determine when the value of  $R(d)$  can be considered significantly close to or distant from 1.0.

### Clustering Analysis

Instead of setting an arbitrary threshold, unsupervised Hierarchical Agglomerative Clustering is applied for partitioning the values of  $R(d)$ . The dendrogram produced from this clustering is presented in Figure 3. In the dendrogram, three major clusters/groups are identified, two marked (by the authors) with dotted rectangles and the third left unmarked in the middle. This middle cluster, denoted as  $G_b$ , represents the set of those developers, who equally express positive and negative emotions during bug-fixing. For this cluster the range of  $R(d)$  is,  $0.922 \leq R(d) \leq 1.0, \forall d \in G_b$ .

The set of developers who are included in the right-most cluster exhibit more positive emotions compared to negative emotions during bug-fixing. Let  $G_p$  denote the cluster of these developers. Here the range of  $R(d)$  is,  $1.005 \leq R(d) \leq 1.178, \forall d \in G_p$ .

The set of developers who render more negative emotions towards bug-fixing belong to left-most cluster, denoted as  $G_n$ . Again, for this cluster the range of  $R(d)$  is,  $0.919 \leq R(d) \leq 0.982, \forall d \in G_n$ .

### Statistical Significance

For each of the three clusters, *MWW* tests are separately conducted between the mean positive and negative emotional scores of the commit messages to verify the statistical significances of their differences. The results of the separate *MWW* tests (with  $\alpha = 0.05$ ) over each of the clusters are presented in Table 5. The *P*-values in Table 5 indicate statistical significance in the differences in

Figure 3. Hierarchical agglomerative clustering of 30 developers enumerated as 1,2,3,...,30

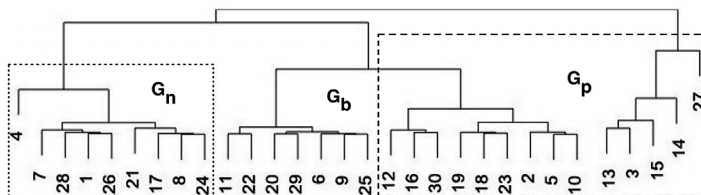


Table 5. *MWW* tests over  $R(d)$  scores of commit messages written by developers in each cluster

| Cluster                    | $G_p$                | $G_n$                | $G_b$               |
|----------------------------|----------------------|----------------------|---------------------|
| <i>P</i> -values           | 0.00798              | 0.0268               | 0.26109             |
| Statistically Significant? | Yes ( $P < \alpha$ ) | Yes ( $p < \alpha$ ) | No ( $p > \alpha$ ) |

positive and negative emotions for clusters  $G_p$  and  $G_n$ . As expected, no such significant difference found for the cluster  $G_b$  as in this cluster, positive and negative emotions are expressed equally. Thus, clustering of the developers appears to be accurate with statistical significance. Hence, answer of the research question RQ2 is formulated as in Box 2.

### Emotional Variations in Days and Times

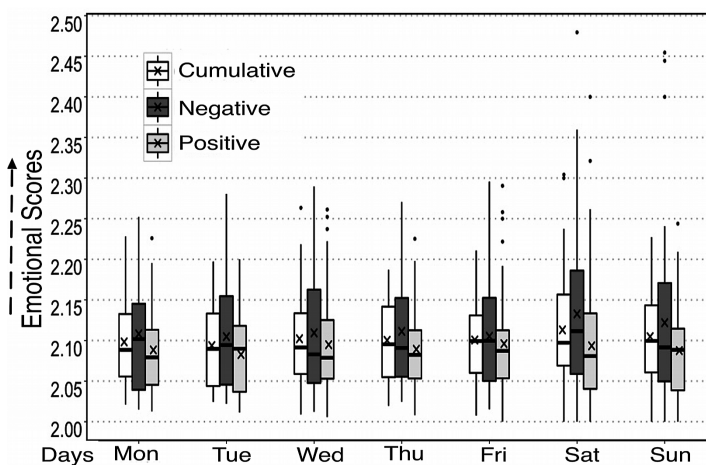
For each of the projects, all the commit messages are grouped into seven disjoint sets in accordance with the days of the week those are committed.

Figure 4 plots the average (over each project) positive, negative, and cumulative emotional scores in commit messages posted in different days in a week. Among all the seven days of a week, negative emotions appear to be slightly higher in commit messages posted during the weekends (i.e., Saturday and Sunday) than those posted in weekdays (i.e., Monday through Friday). Not much difference is visible in the emotional scores for commit messages posted in the five weekdays. *MWW* tests (with

#### Box 2

**Ans. to RQ2:** It is possible to distinguish developers who show either high positive or high negative emotions in bug-fixing commit messages while some other developers are found to express both positive and negative emotions almost equally. The same approach can be applied to distinguish such groups of developers for other types of development tasks.

Figure 4. Distribution of mean positive, negative, and cumulative emotional scores in commit comments posted in different days of week



$\alpha = 0.05$ ) between the distributions of emotional scores in each possible pair of the days of a week suggest no statistical significance in the differences of emotions.  $P$ -values of the  $MWW$  tests are presented in Table 6. As it can be seen in Table 6, for all values of  $P$ 's,  $\alpha < 0.11 \leq P$ .

To study the relationship between developers' emotions and times of a day when commit comments are posted, the 24 hours of a day are divided in three periods (a) 00 to 08 hours as before working hours, (b) 09 to 17 hours as regular working hours and (c) 18 to 23 hours as after working hours. Then for each project, all the commit messages are again organized into three disjoint sets based on their timestamps of posting.

Figure 5 presents the mean (over each project) positive and negative emotional scores (computed using Equation 1 and Equation 2) in commit messages posted in these three periods. Again, in Figure 5, not much variation is seen in the emotional scores of commit messages posted at different periods.  $MWW$  tests (with  $\alpha = 0.05$ ) between the distributions of mean positive and negative emotional scores in each possible pair of the periods indicate no statistical significance in their differences.  $P$ -values from the  $MWW$  tests are presented in Table 7. Hence, the answer to the research question RQ3 is given as in Box 3.

Table 6. Results of MWW tests (P-values) over cumulative emotional scores of commit messages written in different days of week

| Day | Sat  | Sun  | Mon  | Tue  | Wed  | Thu  | Fri  |
|-----|------|------|------|------|------|------|------|
| Sat | -    | 0.44 | 0.23 | 0.11 | 0.33 | 0.41 | 0.35 |
| Sun | 0.44 | -    | 0.71 | 0.42 | 0.77 | 0.98 | 0.84 |
| Mon | 0.23 | 0.71 | -    | 0.68 | 0.79 | 0.71 | 0.84 |
| Tue | 0.11 | 0.42 | 0.68 | -    | 0.55 | 0.41 | 0.49 |
| Wed | 0.33 | 0.77 | 0.79 | 0.55 | -    | 0.83 | 0.96 |
| Thu | 0.41 | 0.98 | 0.71 | 0.41 | 0.83 | -    | 0.86 |
| Fri | 0.35 | 0.84 | 0.84 | 0.49 | 0.96 | 0.86 | -    |

Figure 5. Distribution of mean positive, negative, and cumulative emotional scores in commit comments posted in different periods of day

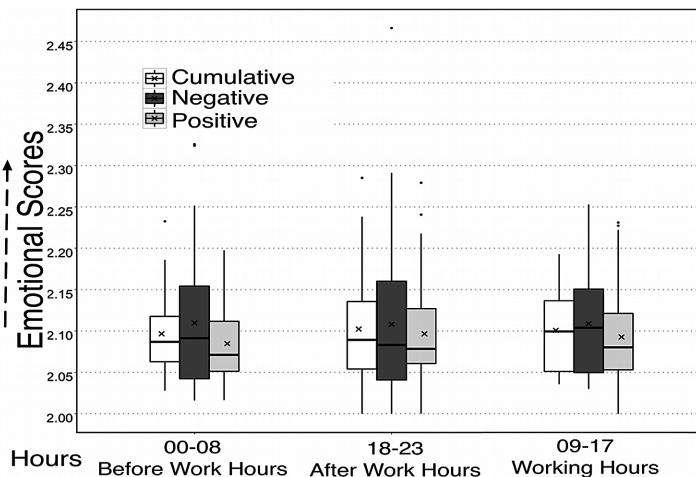


Table 7. Result of MWW tests (P-values) over cumulative emotional scores of commit messages written in different times of a day

| Hours in a Day |       | 00-08   | 09-17   | 18-23   |
|----------------|-------|---------|---------|---------|
| Groups         | 00-08 | -       | 0.59612 | 0.84148 |
|                | 09-17 | 0.59612 | -       | 0.85716 |
|                | 18-23 | 0.84148 | 0.85716 | -       |

Box 3

*Ans. to RQ3 : There are no statistically significant variations in the developers' emotions in different times and days of a week.*

### Emotional Impacts on Commit Lengths

To investigate the existence of any relationship between emotions and lengths of commit messages, across all the 50 projects, a total of 141,033 commit comments are distinguished, which are one to 50 words in length having cumulative emotional scores (computed using Equation 3) higher than one. For each project, these emotional commit messages are organized into four disjoint groups based on their lengths as shown in Figure 6, which plots the mean (over each project) cumulative emotional scores of commit messages in the four groups. As seen in the figure, the emotional scores are strictly higher for the groups with lengthier commit messages.

Table 8 presents the frequencies of commit messages in the four groups of commit lengths and having different cumulative emotional scores between 02 through 08. A Chi-squared (Ramsey & Schafer, 2002) test ( $P = 2.2 \times 10^{-16}, \alpha = 0.05$ ) also strongly indicates statistical significance of the relationship between emotional scores and commits lengths. Next, the significance of the direction of relationship (i.e., if one increases or decreases with the increase of another) is verified.

Figure 6. Distribution of mean cumulative emotional scores of commits comments in groups of different lengths

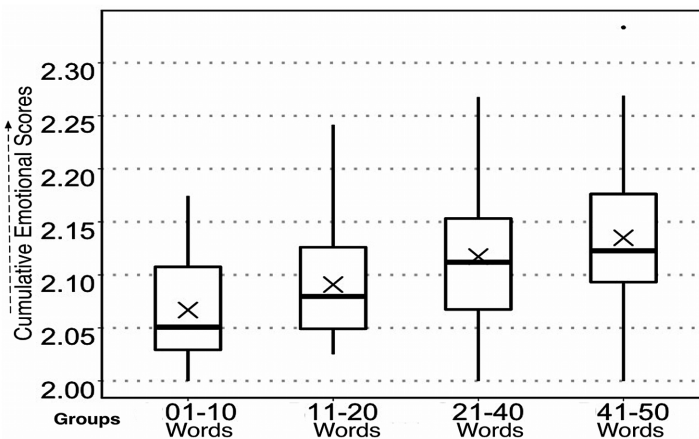


Table 8. Number of commit messages with different lengths (in words) and cumulative emotional scores

| Commit Length |       | Number of Commit Comments with $T(c) =$ |       |       |       |     |    |    |
|---------------|-------|---|-------|-------|-------|-----|----|----|
|               |       | 02                                      | 03    | 04    | 05    | 06  | 07 | 08 |
| Groups        | 01-10 | 46,486                                  | 2,734 | 2,558 | 245   | 28  | 12 | 3  |
|               | 11-20 | 42,144                                  | 3,967 | 4,627 | 876   | 145 | 16 | 1  |
|               | 21-40 | 22,732                                  | 2,633 | 5,008 | 1,155 | 203 | 31 | 2  |
|               | 41-50 | 3,255                                   | 409   | 1,275 | 399   | 84  | 5  | 0  |

Fitting of a Generalized Linear Model (Ramsey & Schafer, 2002) on the emotional score and length of every emotional commit message confirms (with  $\beta = +0.01134, P = 2 \times 10^{-16}, \alpha = 0.001$ ) the positive correlation between emotional scores and commit lengths. Based on the analyses, the answer to the research question RQ4 is derived as in Box 4.

### Impacts of the Sizes of Projects and Teams on Developers' Emotions

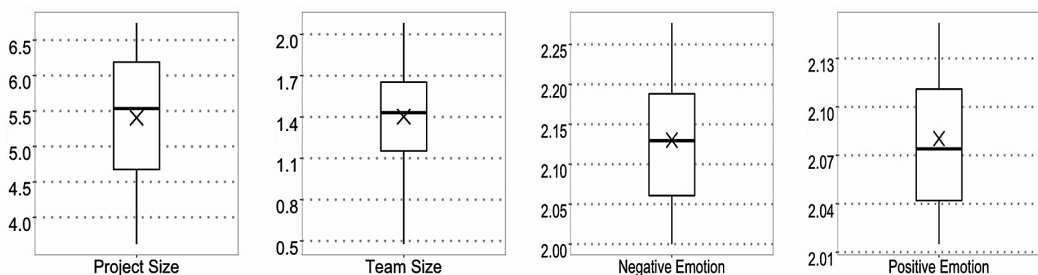
To examine the influence of the sizes of projects and development teams on the developers' emotional variations, the size of each of the 50 projects are extracted by invoking Boa's API. Boa's API provides a project's size in terms of the number of nodes in the AST (Abstract Syntax Tree) constructed by parsing the source code of the project. Intuitively, the number of AST nodes is positively proportional to the number of LOC (Lines of Code). The size of the development team for each of the 50 projects is also determined by computing the number of distinct developers contributing to the project. For each of the projects, the average positive and negative emotional scores per commit message are calculated separately by using Equation 1 and Equation 2.

Figure 7 presents the distribution of sizes of the 50 projects, their team's sizes, and project-wise average negative and positive emotional scores per commit. Notice that all the four distributions in Figure 7 exhibit normal distributions without much skewedness.

#### Box 4

**Ans. to RQ4:** Developers' emotions have statistically significant impacts on the lengths of commit messages they write. Developers post longer commit comments when they are emotionally active.

Figure 7. Distribution of the sizes of the projects, project-wise team-sizes, and mean negative and positive emotional scores over all the 50 projects



For each of the 50 projects, Figure 8 plots the project's size, team's size, as well as the average positive and negative emotional scores per commit message in the project. Note that, the ranges of teams' sizes (198 to 5,871), projects' sizes (4,189 to 5,527,415) and emotional scores (2.00 to 2.36) differ largely. Therefore, logarithm (base 10) operations are applied to the numbers of developers and projects' sizes to properly scale the values for the purpose of legible presentation.

Observing the normal distribution of the data in Figure 7 and the linearity of variations in Figure 8, the *Pearson Product-Moment Correlation Coefficient* (Montgomery, Jennings, & Kulahci, 2008) is applied to examine if the developers emotions are linearly correlated with project's size or team's size. Pearson coefficient is a well-established statistical correlational measurement to examine linear relationship between variables of interval data (Montgomery et al., 2008). For any two variables under consideration, the value of Pearson correlation coefficient ranges between +1.0 and -1.0 and indicates to what extent the variables are positively or negatively correlated. Two variables are positively correlated if one increases, then the other also increases. Negative correlation between variables implies if one gets larger, then the other gets smaller. The closer coefficient to  $\pm 1.0$  the stronger the correlation relationship is.

Table 9 presents the values of Pearson correlation coefficients (with P-values in parenthesis) between the distributions of emotional polarities and the sizes of projects and development teams. As observed in Table 9, positive correlation exists between project size and negative emotional score, and consistently, project size is negatively correlated with positive emotions in developers. These indicate that the developers express more positive emotions while working at smaller projects, perhaps because smaller projects are likely to be easier to deal with.

It is interesting to find (in Table 9) that the team's size is negatively correlated to both positive and negative emotions of the developers. However, the negative correlation of the team's size with negative emotions is relatively stronger (-0.28) while the negative correlation with positive emotions

Figure 8. The numbers of developers, project's size, average positive and negative emotional scores per commit for each of the 50 projects enumerated as 1, 2, 3, ..., 50

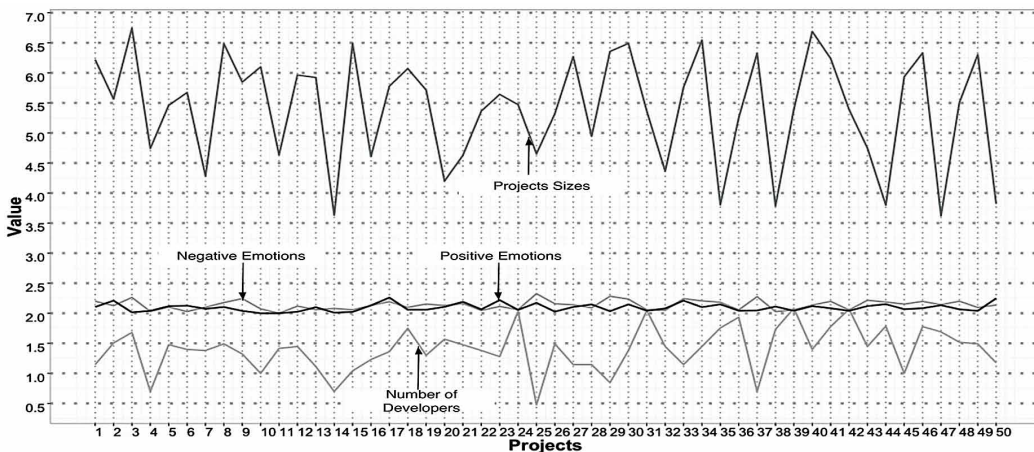


Table 9. Pearson Correlation Coefficients between distributions of emotional polarities and sizes of projects and teams

| Emotional Polarity | Project Size |            | Team Size  |            |
|--------------------|--------------|------------|------------|------------|
|                    | Negative     | +0.27      | (P = 0.05) | -0.28      |
| Positive           | -0.23        | (P = 0.10) | -0.06      | (P = 0.66) |

is close to negligible (-0.06). This can be explained by the possibility that working in larger teams adds communication overhead and collaboration challenges, which might render higher negative emotions in the developers. Indeed, the P-values indicate the observed correlations are not very strong, and thus larger scale studies are necessary to further validate the statistical significance of the discovered correlations.

Based on the observations and statistical tests, the answer to the research question RQ5 is formulated as in Box 5.

## THREATS TO VALIDITY

In this section, the limitations of the work, the threats to the validity of findings, and the authors' attempts to minimize those threats are discussed.

### Internal Validity

The internal validity of this work depends on the accuracy of the tool's computation of emotional scores. *SentiStrength* was reported to be effective in sentiment analysis (Thelwall et al., 2012) and suitable for extraction of emotions from commit comments (Guzman et al., 2014). *SentiStrength* has relatively high accuracy compared to other tools of its kind and thus *SentiStrength* was used for sentiment analysis in earlier work in software engineering research (Chowdhury & Hindle, 2016; Garcia et al., 2013; Guzman et al., 2014; Guzman & Bruegge, 2013; Tourani et al., 2014). Moreover, in this particular work, the accuracy of the sentiment analysis tool was increased for emotion extraction by 26% through tuning for application in software engineering context.

Nevertheless, the tuned tool is not 100% accurate in determining emotional polarities of commit messages, and it was not possible to perform manual sanity check by going through each of the 490,659 commit messages included in this work. The authors are aware of this threat, although they minimized it by contextual customization of *SentiStrength*.

### Construct Validity

One might question the choice of the 30 developers in examining the relationships between emotions and bug-fixing tasks can be questioned. Note that, these 30 developers are the authors of more than 112 thousand commit messages (22.85%), which is a large sample of data for dependable analysis. The objective is to check if it was possible to distinguish a group of developers who are emotionally more active towards a particular type of task. If a fair number of developers are chosen other than the choice of 30, using the approaches of this work, it would still be possible to distinguish a set of target developers. In that case, the size of the set of developers might be different from what was found using the 30 developers, but this does not invalidate the findings of the work.

One may also question the validity of categorization of the developers' commits in different days and periods, considering the possibility that the projects and developers may be physically located at different geographic locations and time-zones. However, it was found that most (86%) of the commits are posted in regular weekdays. Moreover, the majority (58%) of the commit messages are written in regular working hours while 31% and 11% are found to have been posted respectively in before and after regular working hours. The proportions of commits at different days and periods suggest correctness of the categorization.

### Box 5

**Ans. to RQ5:** Developers grow more negative emotions in them while they work in larger projects or in larger development teams, although the statistical strengths of such correlations appear to be not very strong.

In the analysis of the emotional impacts on the lengths of commit messages, commit messages longer than 50 words were excluded, because commit messages of larger lengths include copy-pasted content such as, SQL statements and code snippets. Such contents are not directly created or typed by the committer and thus are unlikely to reflect his or her emotions.

For the statistical tests of significance in the variations of different distributions, the Mann-Whitney-Wilcoxon (MWW) test (Ramsey & Schafer, 2002) is used. The *MWW* test is a non-parametric test, which does not require the data to have normal distribution. Since those portions of data do not conform to normal distribution, this particular test suits well for the purpose. Moreover, the significance level  $\alpha$  set to 0.05, which is a widely adopted value for this parameter that enables 95% confidence in the results of the *MWW* tests.

## External Validity

The findings of this work are based on the study on more than 490 thousand commit messages across 50 open-source projects. This large data-set yields high confidence on the generalizability of the results.

## Reliability

The methodology of data collection, analysis, and results are well documented in this paper. The sentiment analysis tool, *SentiStrength* (Thelwall et al., 2012) is freely available online and projects studied in this work are also freely accessible through Boa (Dyer et al., 2013). Hence, it should be possible to replicate this study.

## RELATED WORK

To explore the impacts of emotions on the debugging performance of software developers, Khan et al. (2010) used high-arousal-invoking and low-arousal-invoking movie clips to trigger different levels of emotions in developers before having them perform some debugging tasks. However, they did not employ any measurement to extract and quantify the developer's emotional states, and relied on the assumption that watching those movie clips would induce different levels of emotions in the developers. Lesiuk (2005) recruited 56 software engineers to understand impact of emotions on software design performance. In her work, music was played to arouse developers' positive emotions. The participants self-assessed their emotional states and design performance. Similarly, self-assessment of emotional states was also used in the studies of Wrobel (2013) and Graziotin et al. (2013).

While the human participants themselves can be expected to accurately report their emotional states, such self-assessment based approaches suffer from the possibility that the participants might be uncomfortable in disclosing their negative emotional states. Biometric measurements such as multi-sensor inputs (McDuff, Karlson, Kapoor, Roseway, & Czerwinski, 2012), audio and video processing (Zeng, Roisman, & Huang, 2009) do not suffer from such difficulties but they can be logistically expensive and difficult for regular use at workplace without disrupting the natural workflow of the practitioners. Both the self-assessment based and biometric approaches for identification of emotions are difficult (if not impossible) to apply for geographically distributed teams and for extraction of emotions from software artifacts of already completed parts of projects.

Note that, unlike this particular work presented in this paper, all of the research mentioned above, focused on understanding the overall emotional impacts over human performance and indicated positive correlation between them. In contrast, this work includes a deeper analysis exploring the impacts and scopes for exploitation of emotions extracted from textual software artifacts such as commit messages. Several other studies also identify developers' emotions from textual software artifacts. In such a study Murgia et al. (2014) reported that issue reports, which express positive emotions take less time to be resolved. They used human raters to identify emotions in issue reports, and thus their work is subject to human errors. Unlike theirs, using an automatic tool *SentiStrength*, this work identifies emotions in a significantly larger number of commit messages. The automatic tool, *SentiStrength* was also used



in the studies of Guzman and Bruegge (2013), Tourani et al. (2014), Garcia et al. (2013), Guzman et al. (2014) and in the work of Chowdhury and Hindle (2016). But none of these work tuned the tool before application in software engineering context, as done in this work.

Guzman and Bruegge (2013) identified emotions in collaboration artifacts to relate them with different development topics. In a separate study, using *SentiStrength*, Guzman et al. (2014) extracted emotions expressed in 60,425 commit messages and reported that commit comments written on Mondays tend to have more negative emotions compared to Sunday, Tuesday, and Wednesday. However, from the investigation of the same phenomenon using a substantially larger dataset of 490,659 commit messages, this study does not identify any statistically significant variations of emotions in commit comments posted in different days of a week.

Using a Natural Language Toolkit (NLTK), Pletea et al. (2014) mined developers' emotions from 60,658 commits and 54,892 pull requests for GitHub projects. They analyzed emotional variations in discussions on different topics and reported to have found higher negative emotions in security-related discussions in comparison with other topics. The study of emotions in security-related development tasks is also a part of this study. While their objective, approach as well as source of emotional content and method of emotion extraction were different from this work, the results of this work derived from a deeper analysis over a substantially larger data-set, support the findings of Pletea et al. (2014).

Using *SentiStrength*, Tourani et al. (2014) extracted emotions from emails of both developers and system users. They observed the differences of emotional expressions between developers and users of a system. Using the same tool Garcia et al. (2013) extracted developers' emotions from their email contents to analyze any relationships between developers' emotions and their activities in an open source software projects. Although the studies of Tourani et al. (2014) and Garcia et al. (2013) also used the same sentiment analysis tool that is used in this work, the source of their emotional content are different and the objectives of those work are also orthogonal to this work.

## CONCLUSION

In this paper, a quantitative empirical study is presented on the characteristics and impacts of emotions extracted from developers' commit messages. More than 490 thousand commit comments over 50 open-source projects are studied. Although the majority (65%) of the commit messages are found to be neutral in emotion, surprisingly, positive emotions are found in relatively much smaller portion (13%) of the commit comments than the commits (22%) containing negative emotions.

In this study, it is found that the polarities of the developers' emotions significantly vary depending on the type of tasks they are engaged in. The developers express equally high positive and negative emotions in committing in energy-aware tasks compared to other tasks. With respect to the polarities of commit messages, positive emotions are found to be significantly higher than negative emotions in commits for bug-fixing and refactoring tasks. On the other hand, negative emotions are significantly higher in security-related commits. Surprisingly, the same scenario is found for new feature implementation commits.

Significant positive correlation is found between the lengths of commit messages and the emotions expressed in developers. When the developers remain emotionally active, they tend to write longer commit comments. The developers tend to render in them more positive emotions when they work in smaller projects or in smaller development teams, although the difference is not very large. Surprisingly, no significant variations are found in the developers' emotions in commit messages posted in different times and days of a week.

Based on emotional contents in commit messages, a group of developers can be distinguished who express more positive emotions at bug-fixing commit messages, another group with the opposite trait, and a third group of developers who equally render both positive and negative emotions at bug-fixing activities. Same approach can be applied for other types of tasks to distinguish potential developers for improved tasks assignment.

The findings from this work are validated in the light of statistical significance. Although more experiments can be conducted to verify or confirm the findings, the results from this study significantly advance the understanding of the impacts of emotions in software development activities and artifacts, and the work also exemplify how emotional awareness can be exploited in improving software engineering activities.

For automatic computation of emotional polarities in commit messages, a state-of-the-art tool, *SentiStrength* has been used while alternatives exist. Moreover, before applying the tool is tuned for this work in the context of software engineering. Future works include replicating this study using other tools and subjects to further validate the findings of this study. The authors also have plans to conduct more studies on the impacts of emotions extracted from diverse artifacts including program comments, development forums and email groups.

## REFERENCES

- Ayalew, Y., & Mguni, K. (2013). An assessment of changeability of open source software. *Computer and Information Science*, 6(3), 68–79. doi:10.5539/cis.v6n3p68
- Choudhury, M. D., & Counts, S. (2013). Understanding affect in the workplace via social media. *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, San Antonio, TX, USA (pp 303–316). doi:10.1145/2441776.2441812
- Chowdhury, S. A., & Hindle, A. (2016). Characterizing energy-aware software projects: Are they different? *Proceedings of the 13th International Conference on Mining Software Repository*, Austin, TX, USA (pp 508–511). doi:10.1145/2901739.2903494
- Destefanis, G., Ortu, M., Counsell, S., Marchesi, M., & Tonelli, R. (2015). *Software development: do good manners matter?* (pp. 1–17). PeerJ PrePrints.
- Dewan, P. (2015). Towards emotion-based collaborative software engineering. *Proceedings of the Eighth International Workshop on Cooperative and Human Aspects of Software Engineering*, Florence, Italy (pp 109–112). doi:10.1109/CHASE.2015.32
- Dyer, R., Nguyen, H. A., Rajan, H., & Nguyen, T. N. (2013). Boa: A language and infrastructure for analyzing ultra-large-scale software repositories. *Proceedings of the 2013 International Conference on Software Engineering*, San Francisco, CA, USA (pp 422–431). doi:10.1109/ICSE.2013.6606588
- Feldt, R., Angelis, L., Torkara, R., & Samuelsson, M. (2010). Links between the personalities, views and attitudes of software engineers. *Information and Software Technology*, 52(6), 611–624. doi:10.1016/j.infsof.2010.01.001
- Garcia, D., Zanetti, M. S., & Schweitzer, F. (2013). The role of emotions in contributors activity: A case study on the gentoo community. *Proceedings of the Third International Conference on Cloud and Green Computing*, Karlsruhe, Germany (pp 410–417). doi:10.1109/CGC.2013.71
- Graziotin, D., Wang, X., & Abrahamsson, P. (2013). Are happy developers more productive? the correlation of affective states of software developers and their self-assessed productivity. *Proceedings of the 14th International Conference on Product-Focused, Software Process Improvement*, Paphos, Cyprus (pp 50–64). doi:10.1007/978-3-642-39259-7\_7
- Graziotin, D., Wang, X., & Abrahamsson, P. (2015). Do feelings matter? On the correlation of affects and the self-assessed productivity in software engineering. *Journal of Software Evolution and Process*, 27(7), 467–487. doi:10.1002/smr.1673
- Guzman, E., Azócar, D., & Li Y. (2014). Sentiment analysis of commit comments in github: An empirical study. *Proceedings of the 13th International Conference on Mining Software Repository*, Hyderabad, India (pp 352–355). doi:10.1145/2597073.2597118
- Guzman, E., & Bruegge, B. (2013). Towards emotional awareness in software development teams. *Proceedings of the 2013 9th Joint Meeting on Foundations of Software Engineering*, Saint Petersburg, Russia (pp 671–674). doi:10.1145/2491411.2494578
- Islam, M. R., & Zibran, M. F. (2016). Towards Understanding and Exploiting Developers' Emotional Variations in Software Engineering. *Proceedings of the 14th IEEE/ACIS International Conference on Software Engineering Research, Management and Application*, Towson, Maryland, USA (pp. 185–192).
- Khan, I. A., Brinkman, W., & Hierons, R. (2011). Do moods affect programmers' debug performance? *Cognition Technology and Work*, 13(4), 245–258. doi:10.1007/s10111-010-0164-1
- Lesiuk, T. (2005). The effect of music listening on work performance. *Psychology of Music*, 33(2), 173–191. doi:10.1177/0305735605050650
- Maalej, W., & Happel, H. (2009). From work to word: How do software developers describe their work? *Proceedings of the 6th IEEE International Working Conference on Mining Software Repositories*, Vancouver, BC, Canada (pp. 121–130). doi: 47PM doi:10.1109/MSR.2009.5069490

- Malik, H., Zhao, P., & Godfrey, M. (2015). Going green: An exploratory analysis of energy-related questions. *Proceedings of the 12th Working Conference on Mining Software Repositories*, Florence, Italy (pp 418-421). doi:10.1109/MSR.2015.53
- McDuff, D., Karlson, A., Kapoor, A., Roseway, A., & Czerwinski, M. (2012). Affectaura: an intelligent system for emotional memory. *Proceedings of CHI* (pp. 849–858). doi:10.1145/2207676.2208525
- Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2008). *Introduction to Time Series Analysis and Forecasting*. John Wiley and Sons, Inc.
- Moura, I., Pinto, G., Ebert, F., & Castor, F. (2015). Mining energy-aware commits. *Proceedings of the 12th Working Conference on Mining Software Repositories*, Florence, Italy, (pp. 56-67). doi:10.1109/MSR.2015.13
- Murgia, A., Tourani, P., Adams, B., & Ortu, M. (2014). Do developers feel emotions? an exploratory analysis of emotions in software artifacts. *Proceedings of the 11th Working Conference on Mining Software Repositories*, Hyderabad, India (pp 261-271). doi:10.1145/2597073.2597086
- Palacios, R. L., Ópez, A., Crespo, A., & Acosta, P. (2010). A study of emotions in requirements engineering. *Organizational, Business, and Technological Aspects of the Knowledge Society*, 112, 1–7. doi:10.1007/978-3-642-16324-1\_1
- Pinto, G., Castor, F., & Liu, Y. D. (2014). Mining questions about software energy consumption. *Proceedings of the 11th Working Conference on Mining Software Repositories*, New York, NY, USA (pp 22-31). doi:10.1145/2597073.2597110
- Pletea, D., Vasilescu, B., & Serebrenik, A. (2014). Security and emotion: Sentiment analysis of security discussions on github. *Proceedings of the 11th Working Conference on Mining Software Repositories*, New York, NY, USA (pp 348-351). doi:10.1145/2597073.2597117
- Ramsey, F., & Schafer, D. (2002). *The Statistical Sleuth* (2nd ed.). Duxbury-Thomson Learning.
- Thelwall, M., Buckley, K., & Paltoglou, G. (2012). *Sentiment strength detection for the social web*. *Journal of the American Society for Info. Sciences et Techniques (Paris)*, 63(1), 163–173.
- Tourani, P., Jiang, Y., & Adams, B. (2014). Monitoring sentiment in open source mailing lists – exploratory study on the apache ecosystem. *Proceedings of 24th Annual International Conference on Computer Science and Software Engineering*, Markham, Ontario, Canada (pp 34-44).
- Wrobel, M. R. (2013). Emotions in the software development process. *Proceedings of the 6th International Conference on Human System Interactions*, Sopot (pp 518–523). doi:10.1109/HSI.2013.6577875
- Zeng, Z., Roisman, G. I., & Huang, T. S. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1), 39–58. doi:10.1109/TPAMI.2008.52 PMID:19029545

*Md Rakibul Islam has been a PhD student in Computer Science Department of University of New Orleans since August, 2015. He has completed his BSc from Khulna University, Bangladesh in 2008. Before coming back to academy, he acquired 7 years of extensive practical experiences in software and telecom industries.*

*Minhaz F. Zibran is an Assistant Professor at the Department of Computer Science, University of New Orleans, USA. His research interests include various aspects of software engineering with particular focus on the application of source code analysis and manipulation for the detection of code smells including code clones and their analysis and management in evolving software systems. Minhaz has co-authored many scholarly articles published in ACM and IEEE sponsored international conferences and reputed journals. Throughout his career, Minhaz also earned both teaching and industry experience. He has been actively involved in organizing international conferences (e.g., IWSC'2015, ICPC'2011, SCAM'2012, ICPC'2012, WCRE'2012, ICSM'2013) in his area of research. He is a professional member of the IEEE Computer Society and ACM SIGSOFT.*