What Happens When You Don't Play Nice? Some Effects of Emotions in Asynchronous Communications

Marc J. Lanovaz[1]
[1]*Département de génie informatique et génie logiciel, École Polytechnique de Montréal, Québec, Canada*
marc.lanovaz@umontreal.ca

***Abstract*—The purpose of my project is to examine how perceived emotions affect responding in asynchronous online communications. To examine this issue, I will extract all the messages sent to the R-help mailing list in the past 10 years and measure perceived emotions in questions and responses using the Senti4SD tool. The tool will be used to record the type of emotion conveyed by each message (i.e., positive, neutral, or negative). Then, my analyses will examine the effects of emotion type on the delay of responding, the probability of sending other messages to the list, the length of messages, and the number of messages in the thread. I will use non-parametric tests to uncover associations between these variables.**

I. Problem Statement & Link with Course

Programmers and developers often use online asynchronous communication (e.g., message boards, mailing lists) to ask question to their peers [1], [2]. These asynchronous exchanges may range from cordial to condescending and even become hateful [3]. Examining how negative and positive emotions affect the behavior of others appears important to foster positive working communities. In recent years, researchers have developed tools to identify emotions in text-based communication. One such tool is Senti4SD, which can be used to identify various emotions conveyed in messages [4]. My project proposes to mine the R-help mailing list to examine how the emotions conveyed in a message influence responding. As such, the project is directly linked to the course as I will apply data mining procedures to the R-help mailing list archives as well as statistical analyses to examine how emotions may affect behavior.

II. Research Questions & Motivation

The main purpose of my project is to examine how perceived emotions affect responding in asynchronous online communications. Specifically, my research questions are as follows:

RQ1: *What are the effects of perceived emotion on the delay of responding by others?*

RQ2: *What are the effects of positive and negative emotions on future messaging?*

RQ3: *What are the effects of emotions on message length and number of messages in a thread?*

III. Data Set & Analyses

I will mine the R-help mailing list archives [5] for data covering a ten-year period (i.e., November 2008 to October 2018). For each thread initiated during this time period, I will record the content of the original message and of all its associated responses, the email addresses of all individuals who sent messages and the timestamp for each message. The email addresses will be converted to unique identifiers to facilitate analyses. The time of each original message will be set to zero and the timestamp of each of the responses will be converted so that it represents the time elapsed since the original message.

Then, I will use the Senti4SD tool to identify emotions conveyed in each message. Each message will be categorized in one of three groups: positive, negative, and neutral [4]. I will then use these groupings to respond to each of my research questions. For the first research question, I will apply a Kruskal-Wallis test in which the dependent variable will be the time elapsed since the last message and the grouping variable will be the type of emotion conveyed by the prior message. For the second research question, a chi-square test will be used to compare the probability of a user sending another question following the three types of exchanges. The third research question will involve conducting another Kruskal-Wallis test to examine whether the length and number of messages in a thread are related to the emotions conveyed. To control for familywise error rate related to conducting multiple statistical tests, I will set the alpha at .01 for each of the tests.

IV. Two Related Papers

In 2014, Murgia et al. [3] presented a study exploring emotions in the Apache Software Foundation issue tracking system. Their main conclusion was that developers do express emotions in their asynchronous written communications when working on software development projects, the most common emotions being joy, gratitude and sadness. However, the study relied on human raters to detect emotions, which limited the scope of the analysis that could be conducted. In a more recent study, Calefato et al. [4] described the development and validation of a new tool, Senti4SD, to automatically detect emotions. Their initial validation showed that their tool produced strong agreement with human ratings. My project aims to extend the study conducted by Murgia et al. by using Senti4SD to examine the impact emotions on online asynchronous communication.

V. Time Planning of Project

My goal is to submit the paper by December 21st so that I can spend time with my family during the holiday break. I plan to download the data from the R-help mailing list archives website in the week of October 29. During the weeks

of November 5 and 12, I will extract the data that I will need for my analysis directly from the .mbox files using a Python script. I am out-of-town the following week and will resume work in the week of November 26 during which I will write the R code to conduct my analyses. Writing code should take me a couple of weeks. The final two weeks (i.e., December 10 and 17) will be reserved for graphing the results and writing up my report for submission by the deadline that I have set for myself.

REFERENCES

[1] P. C. Rigby and A. E. Hassan, "What can oss mailing lists tell us? a preliminary psychometric text analysis of the apache developer mailing list," in *Proceedings of the fourth international workshop on mining software repositories*. IEEE Computer Society, 2007, pp. 23–30.

[2] B. Vasilescu, A. Serebrenik, P. Devanbu, and V. Filkov, "How social q&a sites are changing knowledge sharing in open source software communities," in *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM, 2014, pp. 342–354.

[3] A. Murgia, P. Tourani, B. Adams, and M. Ortu, "Do developers feel emotions? an exploratory analysis of emotions in software artifacts," in *Proceedings of the 11th working conference on mining software repositories*. ACM, 2014, pp. 262–271.

[4] F. Calefato, F. Lanubile, F. Maiorano, and N. Novielli, "Sentiment polarity detection for software development," *Empirical Software Engineering*, vol. 23, no. 3, pp. 1352–1382, 2018.

[5] "The r-help archives," https://stat.ethz.ch/pipermail/r-help/, 2018.