Sentiment Analysis of Developer Comments

Exploring Influencing Factors on Emotional Expressiveness in Software Engineering Projects

Lisa Branz, Ligia Pastrán Reina, Julian Richter, Bastian Waizmann, Prof. Dr. Patricia Brockmann

Abstract

Sentiment Analysis of large amounts of text data opens up entirely new possibilities for exploring the dynamics of team members interacting and working on software engineering projects. In this research a dataset from an incident management system was analyzed in regards to gender ratio of commenters, day of the week and emotional content of comments. A commenter's gender was assessed based on their first name. Sentiment Analysis of the comment content was then performed. In a first step, a possible connection between a commenter's gender and emotion displayed in comments was explored statistically. Female commenters were found to display more overall sentiment in comments across a number of emotion categories. Further, a possible connection between sentiment and day of the week was explored. A One-Way ANOVA and Tukey post-hoc analysis revealed that expression of emotion differed significantly between days of the week across a number of emotion categories. Some gender differences were found in regards to emotional expression throughout the week. In a second step, machine learning algorithms will be applied in order to find out whether a commenter's gender can be classified based on their emotional expressiveness and to assess whether there are different types of users showing different types and levels of emotional expression.

Data Source

The **JIRA Repository Dataset**: Understanding Social Aspects of Software Development

700.000	Issues
2 Million	Comments
100.000	Users
4	Open-source projects

The Apache Software Foundation | **Spring** | **JBoss** | **Codehaus**

Research Questions

- ☑ Do male and female commenters differ in emotional expression?
- ☑ Does emotional expressiveness differ between days of the week?
- ☑ Is there an interaction between gender and emotional expressiveness across days of the week?
- ➡ Can a commenter's gender be classified based on emotional expressiveness?
- ➡ Can commenters be clustered based on emotional expressiveness?

Approach

Gender prediction
Predict developer gender from first name using Gender API

→

Database-level work
Select suitable comments for analysis (n = 1.6 million, 97% male) based on accuracy of gender prediction

→

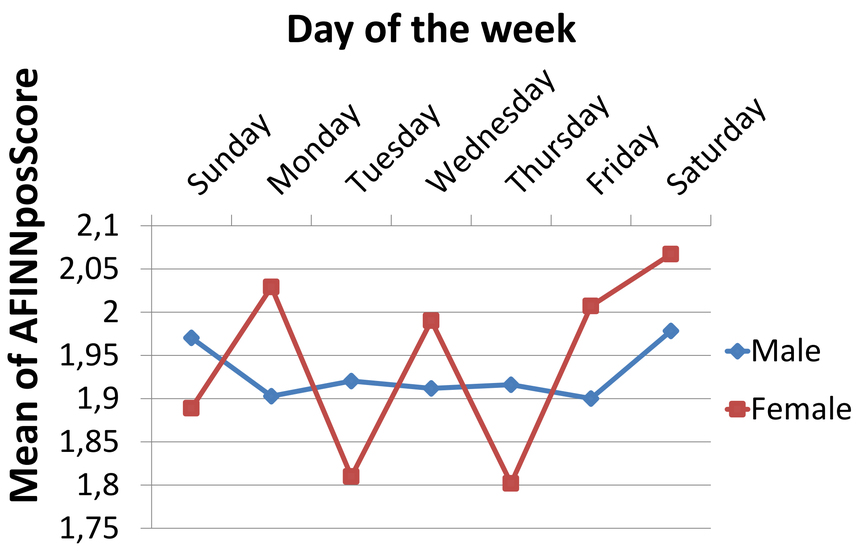
Sentiment Analysis
Classify emotion categories and intensity using the WEKA AffectiveTweets package

→

Statistical Analysis
Explore connections using independent sample t-test, One-Way ANOVA and Tukey post-hoc analysis in IBM SPSS Statistics

Current Results

Males and females differ significantly in most sentiment categories. Females achieved **higher sentiment scores** in most categories. A One-Way ANOVA and Tukey post-hoc analysis revealed that **expression of emotion differed significantly between days of the week** across a number of emotion categories. Gender differences were found in regards to emotional expression throughout the week. While for males mostly Saturday differs significantly from the other days of the week, Monday differs most strongly from other days of the week for females.



Future Directions

- Build a classifier for labeling a comment's author as male or female using emotion scores and days of the week
- Identify the features that add most predictive value when it comes to classifying a commenter's gender
- Find clusters of commenters based on emotion scores in order to assess whether there are different types of commenters showing different types and levels of emotional expression

Goal	Explore how emotional expression of team members in software development teams could influence a team's success, efficiency and team member's satisfaction

TECHNISCHE HOCHSCHULE NÜRNBERG
GEORG SIMON OHM

branzli56977@th-nuernberg.de pastranreinali71943@th-nuernberg.de
richterju71034@th-nuernberg.de waizmannba72103@th-nuernberg.de
patricia.brockmann@th-nuernberg.de

A2IC 2018
ARTIFICIAL INTELLIGENCE INTERNATIONAL CONFERENCE
NOVEMBER 21-23, 2018 | BARCELONA