

A REVIEW OF RESEARCH ON EMOTIONAL STATE ANALYSIS IN SOFTWARE DEVELOPMENT

Simon Park

Department of Electrical Engineering, University of Hawaii, Honolulu, HI, USA.

Abstract: Software development is a complex task. In recent years, emotions in software development activities have received more and more attention. On the one hand, researchers focus on using text sentiment analysis to obtain emotional states in software development, study how to improve the analysis accuracy of existing sentiment analysis tools in the field of software development, and propose effective sentiment analysis tools or methods for the field of software development; on the other hand, researchers focus on using text sentiment analysis to obtain emotional states in software development. On the other hand, researchers have also conducted empirical research on the emotional state and its influencing factors and effects in software development through actual project data. This article explains the emotional state in software development in recent years from the aspects of the current situation of emotional analysis, emotional characteristics in software development, some mainstream tools and main methods of emotional analysis, existing empirical research, and research on the impact of emotion on software development. Current research status of analysis. Finally, some existing problems and future development trends of sentiment analysis in software development are prospected.

Keywords: Sentiment analysis; Opinion mining; Software engineering; Open source community

1 INTRODUCTION

Many current studies have confirmed that in the field of software development, developers can experience various emotions [1-4]. Emotions and emotions are the results of complex psychological activities. Many studies in psychology and neuroscience have shown that emotions are closely related to many cognitive processes. For example, Pessoa's research shows [5] that emotional and cognitive processes use many of the same basic neural circuits, and there are interactions between them. Since software development is essentially a complex cognitive behavior and a social activity, developers need to cooperate with each other and work collaboratively [6]. Therefore, in tasks such as software development, emotions will affect work performance and have a greater impact on productivity, creativity, software quality, team relationships and job satisfaction [7-9]. In recent years, emotions in software development have been There are more and more studies. The academic community holds the SEmotion WorkShop academic seminar every year to discuss research on emotions in software development.

Regarding sentiment analysis in software development, some sentiment analysis tasks currently use survey interview methods [10-13]. These studies are aimed at certain issues in real development activities, such as whether the success of developer meetings will affect The emotions of developers will have an impact on the entire team, as well as the impact of unhappiness and anger on software development. With the popularity of web 2.0, in today's development teams, team members may be distributed in all corners of the world, communicating through various social media, as well as some open source communities such as GitHub and Stack Overflow. The vigorous development of technical Q&A websites has produced a large amount of text information about software development, such as Issue Comments, Commits Comments, posts on technical Q&A websites and online mailing lists. Sentiment analysis of text, or opinion mining, was originally developed to extract the sentiment polarity in short texts published online (such as movie reviews, product reviews, Weibo and Twitter), by analyzing the text to give the current The polarity of the text, positive or negative, thereby trying to unearth some of the thoughts of the creators of these texts [14]. Therefore, many researchers have improved or established many new text sentiment analysis tools or methods according to the characteristics of the software development field, thereby applying text sentiment analysis to the software development field. These studies found that emotions in the open source community do have a certain impact on the quality and efficiency of software development [1, 15-18]. Therefore, extracting an emotional state from these texts for a software development team, a certain piece of software, or a certain software development community can help improve some tools or use some new methods to improve software engineering. Development efficiency, software quality, software maintenance, etc.

Based on the analysis of the current situation of sentiment analysis technology, this article summarizes and analyzes the emotional characteristics in software development, some mainstream tools and main methods of sentiment analysis, existing empirical research, and the impact of emotions on software development. Finally, some existing problems and future development trends of sentiment analysis in software development are prospected.

2 THE CURRENT STATE OF SENTIMENT ANALYSIS

Sentiment analysis of text is currently mainly divided into three levels: text level, sentence level, and clause level. The text level represents the overall emotional tendency of the entire text, the sentence level represents the emotional tendency of a sentence, and the clause level represents the emotional tendency of a phrase or a phrase.

There are two main categories of commonly used sentiment analysis methods: vocabulary-based methods and machine learning or deep learning-based methods. The main tasks of sentiment analysis are: binary classification sentiment polarity analysis, fine-grained sentiment recognition and sentiment intensity analysis.

2.1 Sentiment Analysis Task

For sentiment analysis tasks, sentiment polarity analysis is the basic task of sentiment analysis [19]. Simply put, the text is divided into: positive, negative, and sometimes neutral categories through analysis, but overall it is still a Binary classification method. It has been widely used in the analysis of product reviews and movie reviews. Another method similar to sentiment polarity analysis is sentiment consistency detection, which is a method of analyzing the emotional consistency of two texts, that is, whether they express the same or similar emotions. An improvement to binary classification is to add the intensity of different emotions. After judging the category of emotion, a corresponding score is given based on the intensity of the emotion. This is emotion intensity analysis and an improvement on emotion polarity analysis.

Multimodal fusion sentiment analysis is a newly proposed concept in recent years. Researchers try to use various materials such as videos, voices, pictures, texts, etc., and fuse features extracted from different materials to conduct comprehensive sentiment analysis.

2.2 Vocabulary-Based Approach

Lexicon-based sentiment analysis is a very popular and easy-to-use method with many advantages [20]. First, a pre-annotated sentiment dictionary is required, which contains words such as "happy", "fea", and "sad"., different words give different emotional labels, positive emotions or negative emotions. It can also identify some adverbs of degree such as (very), or some negative structures, such as not happy. These features are combined to identify the emotions in the text. Of course, this model also has some shortcomings. It will produce errors for some grammatically complex phrase recognition. Secondly, this method requires a reliable and consistent emotional dictionary [20], and the language of human society is very complex and contains a lot of comprehensive information. It is difficult for an emotional dictionary to cover many different aspects, so for thinking It is very difficult to dig out the semantic information. Moreover, because the corpus source determines the lexical affinity probability, it is often biased towards specific types of texts. This makes it difficult to develop reusable, domain-independent generalization models [21].

2.3 Machine Learning Based Methods

Sentiment analysis methods based on machine learning are mainly divided into two parts: based on traditional machine learning algorithms and based on deep learning. While traditional machine learning algorithms are usually based on supervised learning, algorithms based on deep learning use some semi-supervised or weakly supervised methods.

The main process of machine learning sentiment analysis is: after the text preprocessing is completed, the text is vectorized to convert the text into a vector that can be used for calculation. The main methods include bag-of-word model, TF-IDF, word2vec, etc. The bag-of-words model, also known as One-hot representation, is an algorithm that counts the number of times a word appears in a document; TF-IDF is a statistical method used to evaluate the relevance of words to a document set or a corpus. The importance of one of the documents; word2vec is a shallow neural network used to output word vectors. After obtaining the numerical vector of the text, use machine learning methods to train the classifier. Some mainstream classification methods include Naive Bayes (NB), Support Vector Machine (SVM), Maximum Entropy (ME), RFC, GNB, Decision Tree, Logistic Regression, etc. The prerequisite for carrying out these machine learning is that a large annotated training text is needed to train the model to ensure accuracy, which has high requirements for preparatory work. In addition, the disadvantage of machine learning learning methods is that the understanding of semantics is somewhat poor, so the classification of sentences or phrases is not very accurate.

With the popularity of deep learning in the field of image processing, deep learning also has good performance in sentiment analysis [22]. This method is similar to machine learning. It also requires some feature extraction methods to vectorize the text. The difference is that the finally extracted features are trained through some neural networks instead of machine learning algorithms. The advantage of deep learning is that these algorithms do not require a lot of manually labeled data sets. They will automatically learn complex features in the data set, but they require a huge amount of data to achieve good performance [23]. Some current research has proven that deep learning has improved to a certain extent compared with traditional methods in sentiment analysis, especially in Weibo, product reviews, and movie reviews [24]. Currently, the mainstream in sentiment analysis uses recurrent neural networks, such as: RNN, LSTM, GRU, etc. There are also some studies using convolutional neural networks CNN, and some studies superimpose convolutional neural networks and recurrent neural networks for sentiment analysis. [25].

3 SENTIMENT ANALYSIS TOOLS FOR SOFTWARE ENGINEERING

In addition to using machine learning methods, some current research also uses SentiStrength [26] to perform emotional polarity analysis and scoring [18, 27, 28]. SentiStrength is a dictionary-based tool that can handle short informal texts, including abbreviations, intensifiers and emoticons. The score of each negative emotion word ranges from -2 to -5, which represents its polarity. Likewise, positive emotion words have a score between + 2 and + 5, while neutral words

have a score of ± 1 . By inputting a sentence, it will calculate the scores of the sentiment words in it, and finally output their sum to represent the score of the entire sentence. There are other sentiment analysis tools, such as the Stanford Sentiment Analysis Tool developed by Stanford, NLTK, etc. However, due to the fact that these tools were originally built for short texts such as product reviews, Twitter, and Weibo, and the particularity of texts in various software development communities, compared with social media texts such as Weibo and Twitter, these software-related texts are The developed text will contain more words from technical fields and include some code snippets. These professional words may have different meanings from their usual meanings. Therefore, when analyzing text information in the field of software engineering, the accuracy of some original tools cannot be guaranteed [29, 30] [19, 20]. Therefore, in order to improve the accuracy of analysis, some tools developed or improved specifically for the field of software engineering have emerged, including SentiStrength-SE [31], SentiCR [32], SentiSW [33], Senti4SD [34], EmoTxt[35] etc. Their main information is shown in Table 1.

Table 1 Sentiment analysis tools in software engineering

name	method	data set
SentiStrength-SE[31]	Dictionary based	issue comments in JIRA
SentiCR[32]	Based on machine learning (GBT)	Open Source Software projects review comments
Senti4SD[34]	Based on machine learning (SVM)	Posts on Stack Overflow
SentiSW[33]	Based on machine learning (GBT)	GitHub projects issue comments
EmoTxt[35]	Based on machine learning (SVM)	Jira&Stack Overflow Posts in

Among them, SentiStrength-SE [31] is a vocabulary-based sentiment analysis tool based on SentiStrength, which is mainly used to identify the polarity of sentiment. It uses 5,992 data extracted from Issue Comments of the JIRA issue tracking system, uses manual annotation of emotional polarity to establish a data set for testing, and collects data based on previous research by Islam. [27] A huge data set, including 490k commit messages of 50 open source projects in Github, from which emotional words for the field of software development are extracted. Through some preprocessing and improvement of negative word recognition rate, the accuracy of emotional polarity discrimination is finally improved., improving SentiStrength's shortcomings in the field of software engineering.

Senti4SD [34] is mainly based on machine learning and uses 4423 posts in Stack Overflow. First, preprocess their contents, and finally use uniform sampling to divide the samples into three equal parts, namely positive, negative, and neutral, with a total of 4800 items. Secondly, the author uses StackOverflow The quantity in is 20M A distributed semantic model (DSM) is established for the post data, which mainly uses word2vec to vectorize the words in the text in order to calculate some similarities. And improve the accuracy of sentiment analysis from three aspects: vocabulary-based features, keyword-based features and semantic features. Finally, the author used the SVM method to train this classifier. For the same StackOverflow data, it reduces the misclassification of neutral sentiment into negative sentiment compared to SentiStrength.

SentiCR [32] is an emotion classification tool based on supervised learning, which is mainly used to analyze the emotions in Code review comments. The author manually annotated 2000 review comments as a training set, and used this training set to test 7 existing sentiment analysis tools and found that the results were not ideal. Therefore, the author performed some specific processing on the data set based on some characteristics of the field of software engineering, and compared the effects of different machine learning algorithms such as ADB, DT, GBT, NB, RF, M LPC, SGD, and SVC., and finally found that GBT has the highest precision and recall rate.

SentiSW [33] is an entity-level sentiment analysis tool that mainly uses machine learning methods for sentiment analysis and a rule-based entity recognition. The author selected 231,732 issue comments from 10 popular GitHub projects. 3000 pieces of data were selected for manual annotation as a data set, and an entity-level emotion classification tool was designed. This entity recognition is to distinguish whether the emotional expression in the text is for people or things. First, the noise in the data set was removed through preprocessing, such as removing URLs and expanding abbreviations. Then, TF-IDF and Doc2vec, two text vectorization methods, were used to vectorize the text. The former is a statistical method, the latter is a new version that adds paragraph vectors to Word2vec. Then we compared the precision and recall rates of six machine learning methods horizontally, and found that the stochastic gradient boosting tree GBT performed the best. Finally, rule-based entity recognition is used to identify issues. Is the object of the sentiment in comment a person or an item, and outputs a tuple of (sentiment, object) as the result.

EmoTxt [35] is a classifier trained by a support vector machine (SVM). It uses 4800 posts in StackOverflow as a training set and can distinguish multiple emotions such as Joy, Love, Sadness, Anger, Surprise, and Fear.

In addition to these tools, Daviti. [7] built a classifier to identify anger, because they believed that detecting anger may help support developers who encounter difficulties, contribute to the management of software communities, and contribute to the management of software communities. Recommendations and optimization improvements for projects. In addition, some researchers have conducted comparative evaluations on the above tools. A benchmark study on sentiment analysis tools [36] analyzed and compared SentiStrength, Senti4SD, and SentiStrength-SE. The improved

tools performed better than SentiStrength, which shows that tools developed in the field of software development do improve classification accuracy. When other data sets are used for testing, such as Code Review and Java libraries, although the accuracy decreases, it is still within the acceptable range. However, some studies have shown that most of the above tools cannot meet acceptable accuracy requirements for the data sets they use [37]. Therefore reliable sentiment analysis in software engineering is possible, but there are still many shortcomings.

3.1 Data Sets Related to Sentiment Analysis in the Field of Software Engineering

Currently, the main data sets related to sentiment analysis in the field of software engineering include the JIRA warehouse data set released by Ortu. [38], which mainly obtains Issue Tracking from the Apache Software Foundation, Spring, JBoss and CodeHaus. System relevant data. It contains more than 1,000 projects, as well as more than 700,000 Issue reports and 2 million Issue comments. The author divides the issue into three parts: bugs, upgrades, and feature requests, and stores them all in a SQL-type database.

Later, based on this data set, the author released a data set containing 2,000 issue comments and more than 4,000 sentences, and manually annotated the emotions contained in these sentences by developers, such as love, happiness, surprise, and anger., disappointment, fear, etc. These data are divided into three groups, including 392 comments, 1600 comments, and 4000 sentences respectively, and are composed of 16 One, three, and three volunteers performed data annotation. When labeling 392 comments data, it was found that among all emotions, love, happiness, and sadness have the least disagreements. Therefore, the latter two groups only selected labels from love, happiness, and sadness, and found that the proportion of neutral emotions in the data was very high, reaching 43.4%. The final set of 4,000 sentences was annotated in a more fine-grained manner.

However, since sentiment analysis in the field of software engineering is still in its infancy, and there is no recognized data set for the time being, many studies are based on their own needs, and more data are collected and annotated again to meet the needs of research.

3.2 The Difference between Software Development-Related Text Data and Traditional Sentiment Analysis Data

Traditional social media data, such as Weibo, Twitter, product reviews, etc., are compared with software development-related text. First, as far as the content of the text is concerned, traditional social media data is all-encompassing and has many different topics, while the topics around the text in software development are only about the field of software development. Second, in comparison with the emotional information contained in it, the emotions expressed in data such as product reviews are usually more obvious. As for text data related to software development, some existing studies have found that it contains more neutral emotions. Third, in terms of text length, data such as Weibo is usually very short text (less than 140 words) due to its own limitations, but there is no such restriction in software development, so the length span of the text is relatively long. big. This also has an impact on the choice of sentiment analysis methods. Fourth, due to the particularity of the field of software development, texts usually contain a lot of code and common vocabulary in software development. However, these words are different from their general meanings, and the emotions they express are also very different. Therefore, if Using models built with traditional social media data to predict software development-related texts will lead to a decrease in prediction accuracy.

4 EMPIRICAL RESEARCH ON SENTIMENT ANALYSIS IN THE FIELD OF SOFTWARE ENGINEERING

In terms of the emotional theory of emotional analysis in software development, we mainly refer to two emotional frameworks in psychology. One is the basic emotion theory, which is a discrete framework, that is, emotions are divided into different basic categories such as happiness, sadness, and anger. Emotion, which is also the most widely used one, also has a dimensional theory of emotion, which believes that emotion is continuous rather than discrete in the brain. Many researchers believe that emotion is at least two-dimensional and proposed the valence-arousal model (Valence -Arousal) [39] believed that emotion includes two aspects: valence and arousal. Valence is defined as the degree of happiness or sadness, while arousal is defined as a subjective state of sensory activation or deactivation, and emotion is this A mixture of the two. This is also used by some research studies [3]. Many current studies mostly descriptively analyze the emotional information contained in various textual materials in software development. For example, Commit comments [15, 16, 27] in GitHub, Pull requests[16], Commit logs[17], Code review, issue comments[40], JIRA Issues in comments,Apache mailing in lists[18], and some other IT text data[41]. These studies select different data sets, mainly text information in the open source community, and use some emotional analysis tools and some machine learning text classification algorithms to extract emotional information in the text. Other studies focus on the impact of emotion on software development, such as the efficiency of Issue Fixing, the speed of debugging, whether there is a difference in efficiency between developers with different emotions, and whether there is a connection between a good team atmosphere and the quality of software, etc. wait.

4.1 The Current Situation of Emotion in Software Development and Factors Affecting Emotion

There have been many studies that have done a lot of descriptive analysis on the current status of emotions in software development, from the proportion of emotional polarity in the entire data set, and the degree of expression of positive emotions and negative emotions. Since the foundation of sentiment analysis is based on two different emotional

frameworks in psychology, namely the discrete emotion and the valence-arousal model, the research mainly focuses on classification and adoption of the VAD (Valence, Arousal, Dominance) evaluation model. These two aspects are carried out. In addition to the differences in basic emotion theories, the current study also analyzes the influencing factors of emotion in software development, mainly from different text sources, the identity of the text publisher, the language used in the project, date or time, on different topics Issues, different development tasks (bug fixing, Refactoring, etc.), to explore their impact on emotions.

4.1.1 Character's perspective

From the user's perspective, Ortu. [40] believe that different user types have a certain impact on emotions. They divide the reviewers of Issues in Github into two types, pure reviewers and reviewers who are also contributors. The author calls the former commentators and the latter users. Based on the results of the study, they found that commenters' comments on Issues were less polite and contained more negative emotions than users' comments, and that commenters expressed fewer emotions, while they communicated excitement, emotion, and dominance. The level is higher. This also shows that there are certain differences in the emotions expressed between different user types and different positions.

4.1.2 Theme's perspective

Pletea. [16] focused on the security-related Commits on GitHub. And pull request, from the perspective of topic, analyze the emotional differences between security topics and other topics. They used keyword extraction and clustering methods to extract topics from the text, and used NLTK to analyze the polarity of sentiment. After analyzing the results, the author found that security-related topics showed more negative emotions, or were more emotional, than non-security topics. Islam. [16] According to different task types in software development, such as Bug-fixing, Energy-Aware, New feature, Refactoring, to analyze their impact on emotions, and found that in terms of energy-aware, positive emotions are roughly equivalent to negative emotions, in terms of Bug-fixing and refactoring, positive emotions are much higher than negative emotions, while for New feature, the opposite is true. Different from the study by Lazar. [17], Islam did not find any difference in emotions regarding dates. Finally, it was analyzed that developers tend to post longer comments when they are in a better mood.

4.1.3 Perspective of time

Lazar. [17] extracted commits logs from approximately 28,466 GitHub projects in 7 years. Through SentiStrength analysis, it is found that about 74% of the commits have neutral sentiment, 7.19% are positive, and the remaining 18.05% are negative. However, the author did not perform preprocessing before analysis. The code blocks and URL addresses contained in the text may cause many neutral components to be added to the results. In addition, the author also used the date of submission as a factor that affects emotions, and conducted a comparative analysis from Monday to Sunday, and found that Tuesday has more negative emotions, and for the top five major projects, Wednesday and Thursday have the most negative emotions. And a positive correlation was found between the number of file modifications and sentiment.

4.1.4 Team perspective

An early study by Tourani. [18] analyzed the identification and assessment of pain or happiness in development teams. They selected the mailing lists of two of the most successful and mature projects in the Apache Software Foundation and conducted sentiment analysis on them. They found that user and developer mailing lists elicited both positive and negative sentiment, and that sentiment analysis tools were not as accurate for email messages as they were for tweets. They are relatively long compared to the reviews posted. Guzman. [15] took commit comments in open source projects as a data set and analyzed them from four aspects: the language used by the project, time and date, the distribution of the team, and the recognition of the project. They mainly used the SentiStrength tool to analyze Emotions were scored and analyzed statistically using the Wilcoxon rank sum test. The sentiment scores in these data sets are relatively close to neutral. This may be because the comments are mainly about technology and are not very emotionally rich. Finally, the author concluded that the Java language contains more negative emotions. Comments on Mondays of the week show more negative emotions. At various times of the day, comments in the afternoon are more positive than those in the evening. In addition, research has also found that the more developers involved in a project, the positive emotions contained in comments will also increase. Karl.[42] studied from the perspective of the impact of project duration on emotions as the project progresses, whether the emotions of the entire team will change, as well as the leader's emotions, the number of project participants and the size of the project. Will it have any impact on emotions? The results show that as time continues, the team's positive emotions decline. This trend is slowed down in the middle of the project cycle, but intensifies towards the end of the project. The more positive the leader's emotions show, the smaller the tendency for the project's positive emotions to decrease.

4.2 The Impact of Emotions on Software Development

Emotions have many impacts on human behavior. This has been confirmed in many psychological studies, and software development activities are also affected by emotions. From current research, the impact of emotion on software development is mainly concentrated in the aspects of efficiency, quality, team atmosphere, and developer behavior.

From the perspective of efficiency, Ortu.[4] analyzed the relationship between emotion and issue resolution time, selected Issue comments of some projects from Jira issues, and analyzed them from the aspects of joy, love, sadness, and anger. analyze. The results show that emotion does have a significant impact on the resolution time of issues. The more positive emotions, the shorter the resolution time of issues. The more negative emotions, the longer the resolution

time of issues. Ortu. [43] in another study 14 open source software projects were selected for analysis and found that politeness also has an impact on the resolution time of Issues. The resolution time of polite Issues is lower than that of impolite Issues; there is a positive relationship between politeness and the attractiveness of the project and the stickiness of developers. Relatedly, items expressed more politely in issues usually leave developers with positive emotions and show greater appeal. Graziotin.[44] investigated the relationship between software developers' emotional state, creativity and analytical problem-solving skills. Starting from the theoretical perspective of the emotional dimension of VAD, the researcher used two questionnaires, SAM (Self-Affective Assessment Questionnaire) and Self-Work Productivity Assessment Questionnaire, to test developers, thereby obtaining developers' performance in Valence, Arousal, and Dominance. scores and productivity scores. After conducting data analysis, it was found that the emotional state of software developers was positively correlated with their self-assessed productivity.

From the perspective of team atmosphere, Garcia. [45] used data from bug reports and mailing lists of the GENTOO community (a Linux community) to analyze the role of emotion in developer activity. The author first investigated the emotional factors related to the departure of a core developer, Alice, and found that the content of emails between her and the developer was more negative than other people's, and that after Alice left, it had a lot of negative effects on the community. Then it was further extended to all developers, and the relationship between the developer's emotional expression and the time interval of individual inactivity was analyzed. It was found that after a certain emotional expression occurs, developers will have a decrease in motivation, so that they can Estimate when a developer becomes inactive based on the sentiment he expressed in his last message. There are also studies that believe that emotional trust has an impact on distributed development teams. Trust is the basis of software projects, especially in distributed situations. At the same time, emotional trust is considered to be a predictor of successful distributed software collaboration. It is speculated that in distributed projects, the greater the emotional vocabulary in the previous pull-request annotation, the greater the chance that the current pull request will be accepted, but this has not been confirmed by further experiments [46].

In terms of work performance, or quality, Khan. [47] conducted two different experiments on developers. They asked developers to watch videos to induce specific emotions, then conducted debugging tests, and first conducted Some physical exercises, then doing algorithmic calculation tasks and testing their performance and emotional state. The results show that the impact of different states of emotional valence (Valence) on debugging tasks does exist. After watching low-arousing and high-arousing video clips, the results are also significantly different; after physical exercise, the programmer's arousal and valence were significantly improved, and at the same time, their task performance also improved. Quality control is also very important in the software development process. Francisco Jurado. [48] suggested introducing sentiment analysis technology into the software development process to identify and monitor text written by developers in issues and tickets. underlying emotions. Therefore, the author analyzed the data of 9 well-known projects and found that there are indeed potential emotional expressions in these data. Therefore, the author believes that this information can be used to analyze the development process and monitor emotional characteristics during the development process.

5 ANALYSIS AND OUTLOOK

To sum up, it is not difficult to see that sentiment analysis research in the field of software development has achieved certain results. The current situation of emotions in software development has been analyzed from many angles, but there are still some directions that can be explored. At present, many researchers have developed sentiment analysis tools customized for the field of software engineering, and some studies have analyzed the impact of sentiment on software development. These tools have good performance on some data sets, but some studies have pointed out that the performance when tested on other data sets is much lower than expected, which shows that the current sentiment analysis tools or methods in the field of software engineering There is still a lot of room for improvement. Therefore, future research directions are mainly elaborated from the following aspects.

5.1 Challenges Faced by Sentiment Analysis in Software Engineering

At present, most analyzes of emotional states in software development adopt polar classification methods. However, some studies believe [49] that since emotions are very complex, a single emotional polarity is not enough to dig out developers' emotions, which is only one dimension of emotion. In addition, some existing sentiment analysis tools in the field of software development have good internal validity, but relatively poor external validity. That is to say, although they are all texts related to software development, they are not The performance of tools developed on StackOverflow data in Github data may not reach its original level [34]. Also, when using machine learning methods to train sentiment analysis models, you may encounter situations where overfitting causes good performance on the training set but poor performance on the test set. These are the difficulties that need to be overcome in sentiment analysis in the field of software development.

5.2 Prospects of Research on Sentiment Analysis in Software Development

5.2.1 More accurate sentiment analysis for software development

At present, sentiment analysis methods are mainly based on dictionary and machine learning methods. There are also many sentiment analysis tools for the field of software engineering, but the overall effect still has a lot of room for

improvement. However, deep learning technology has performed well in sentiment analysis in other fields and has been widely used. Recently, Achyudh Ram.[25] established a deep network combining CNN and LSTM to perform sentiment analysis on 5 different data sets about software development. Compared with traditional machine learning algorithms and some dictionary-based sentiment analysis tools, Deep networks do have certain advantages, but the author also points out that there are certain prerequisites for using deep learning technology, such as the size of the data set, the distribution of the content of the data set, etc. Therefore, in the following research, the sentiment analysis method based on deep learning can be applied to software engineering data sets.

5.2.2 Deeper sentiment analysis for software development

When you have more accurate sentiment analysis tools, you can dig out more sources of information and conduct more in-depth sentiment analysis. Some social media text sentiment analysis studies believe that emoji, text emoticons and other information contain emotional information, and analyze them together with the text. In the field of software development, Clase. [50] extracted emoticons from the Issue comments of Apache and Mozilla. They found that developers at Apache and Mozilla mostly use emoticons to express joy, and some express sadness or surprise. This also shows that developers are indeed frustrated. At the same time, the researchers also said that this is only the first step in the study of emoticons. In the future, it can be combined with some sentiment analysis tools or other sentiment analysis technologies. In addition to this kind of text information, multi-modal sentiment analysis is also gradually receiving attention. This is different from traditional text sentiment analysis. It can integrate many information such as audio, video, text, etc. to perform sentiment analysis. In some software development fields or related social media, it may be possible to obtain corresponding audio, video, pictures and other materials to conduct multi-modal sentiment analysis.

5.2.3 Wider sentiment analysis for software development

Finally, we need to analyze sentiment not only from the perspective of each open source project, but also from a broader scope, such as the atmosphere of the entire project community, not just large communities like GitHub, or ITS systems like Jira, There are also some other communities worthy of attention, such as various Ubuntu, Gentoo etc. In addition to such horizontal comparisons, we can also do more longitudinal analysis in the future to try to understand the changes in sentiment in the project over a time span. Ferreira. [51] analyzed the sentiment in the Linux kernel development mailing list over time. Has there been any change in the passage of time? Perhaps we can observe the development and evolution process of a project or a community, and whether there is some connection between this process and emotions.

6 CONCLUSION

The relationship between emotion and software development is very close. At the same time, the rise of the open source community has brought a massive amount of data. Therefore, in recent years, based on these huge amounts of data, researchers have used various sentiment analysis techniques to analyze the software development process. Emotional states have been analyzed. Although great progress has been made in this field, due to the particularity of software development, the effectiveness of emotion analysis tools in software engineering still needs to be improved. On the other hand, emotions have a wide impact on the software development process. A good emotional state is important for improving the efficiency and quality of software development, the atmosphere of the software community, management, communication and collaboration of the software team, and the enthusiasm and stickiness of developers. Can help. Therefore, it is of important research value to explore the factors that affect emotional states and how to use the regulation of emotions to improve the software development process. The current research has just started and needs to be deepened urgently.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

- [1] Murgia A, Tourani P, Adams B. Do developers feel emotions? an exploratory analysis of emotions in software artifacts//Proceedings of the 11th Working Conference on Mining Software Repositories(MSR), ACM, 2014:262-271.
- [2] Girardi D, Lanubile F, Novielli N. Sensing developers' emotions: the design of a replicated experiment//IEEE/ACM 3rd International Workshop on Emotion Awareness in Software Engineering (SEmotion), IEEE, 2018:51-54.
- [3] Mntyl M, Adams B, Destefanis G. Mining valence, arousal, and dominance: possibilities for detecting burnout and productivity?//Proceedings of the 13th International Conference on Mining Software Repositories(MSR), ACM, 2016:247-258.
- [4] Ortu M, Murgia A, Destefanis G. The emotional side of software developers in JIRA//Proceedings of the 13th International Conference on Mining Software Repositories(MSR), ACM, 2016: 480-483.
- [5] Pessoa L. On the relationship between emotion and cognition. *Nature Reviews Neuroscience*, 2008, 9(2): 148-158.
- [6] Storey M A. The evolution of the social programmer//Proceedings of the 9th IEEE Working Conference on Mining Software Repositories(MSR), IEEE Press, 2012: 140-140.

- [7] Gachechiladze D, Lanubile F, Novielli N. Anger and its direction in collaborative software development//IEEE/ACM 39th International Conference on Software Engineering: New Ideas and Emerging Technologies Results Track (ICSE-NIER), IEEE, 2017: 11-14.
- [8] De Choudhury M, Counts S. Understanding affect in the workplace via social media//Proceedings of the 2013 Conference on Computer Supported Cooperative Work, ACM, 2013: 303-316.
- [9] Ortu M, Adams B, Destefanis G. Are bullies more productive?: empirical study of affectiveness vs. issue fixing time//Proceedings of the 12th Working Conference on Mining Software Repositories(MSR), IEEE Press, 2015: 303-313.
- [10] Calikli G, Al-Eryani M, Baldebo E. Effects of automated competency evaluation on software engineers' emotions and motivation: a case study//IEEE/ACM 3rd International Workshop on Emotion Awareness in Software Engineering (SEmotion), IEEE, 2018:44-50.
- [11] Kuuttila M, Mntyl M V, Claes M. Daily questionnaire to assess self-reported well-being during a software development project//Proceedings of the 3rd International Workshop on Emotion Awareness in Software Engineering(SEmotion), ACM, 2018:39-43.
- [12] Graziotin D, Fagerholm F, Wang X. Consequences of unhappiness while developing software//Proceedings of the 2nd International Workshop on Emotion Awareness in Software Engineering(SEmotion), IEEE Press, 2017:42-47.
- [13] Sarkar S, Parnin C. Characterizing and predicting mental fatigue during programming tasks//Proceedings of the 2nd International Workshop on Emotion Awareness in Software Engineering (SEmotion), IEEE Press, 2017: 32-37.
- [14] Pang B, Lee L. Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2008, 2(1-2): 1-135.
- [15] Guzman E, Az6car D, Li Y. Sentiment analysis of commit comments in GitHub: an empirical study//Proceedings of the 11th Working Conference on Mining Software Repositories (MSR), ACM, 2014: 352-355.
- [16] Pletea D, Vasilescu B, Serebrenik A. Security and emotion: sentiment analysis of security discussions on GitHub//Proceedings of the 11th Working Conference on Mining Software Repositories (MSR), ACM, 2014: 348-351.
- [17] Sinha V, Lazar A, Sharif B. Analyzing developer sentiment in commit logs//Proceedings of the 13th International Conference on Mining Software Repositories(MSR), ACM, 2016:520-523.
- [18] Tourani P, Jiang Y, Adams B. 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(CSSE), IBM Corp, 2014:34-44.
- [19] Cambria E. Affective computing and sentiment analysis. IEEE Intelligent Systems, 2016, 31(2): 102-107.
- [20] Taboada M, Brooke J, Tofiloski M. Lexicon-based methods for sentiment analysis. Computational Linguistics, 2011, 37 (2): 267-307.
- [21] Cambria E, Schuller B, Xia Y. New avenues in opinion mining and sentiment analysis. IEEE Intelligent Systems, 2013, 28 (2): 15-21.
- [22] Collobert R, Weston J, Bottou L. Natural language processing (almost) from scratch. Journal of Machine Learning Research, 2011, 12(8): 2493-2537.
- [23] Mikolov T, Chen K, Corrado G. Efficient estimation of word representations in vector space. arXiv preprint arXiv: 1301.3781, 2013.
- [24] Araque O, Corcuera-Platas I, Sanchez-Rada J F. Enhancing deep learning sentiment analysis with ensemble techniques in social applications. Expert Systems with Applications, 2017, 77(C): 236-246.
- [25] Ram A, Nagappan M. Supervised sentiment classification with CNNs for diverse SE datasets. arXiv preprint arXiv: 1812.09653, 2018.
- [26] Thelwall M, Buckley K, Paltoglou G. Sentiment strength detection for the social web. Journal of the American Society for Information Science and Technology, 2012, 63(1): 163-173.
- [27] Islam M R, Zibran M F. Towards understanding and exploiting developers' emotional variations in software engineering//IEEE 14th International Conference on Software Engineering Research, Management and Applications (SERA), IEEE, 2016: 185-192.
- [28] Guzman E, Bruegge B. Towards emotional awareness in software development teams//Proceedings of the 2013 9th Joint Meeting on Foundations of Software Engineering(ESEC/FSE), ACM, 2013:671-674.
- [29] Imtiaz N, Middleton J, Girouard P. Sentiment and politeness analysis tools on developer discussions are unreliable, but so are people//IEEE/ACM 3rd International Workshop on Emotion Awareness in Software Engineering (SEmotion), IEEE, 2018:55-61.
- [30] Jongeling R, Sarkar P, Datta S. On negative results when using sentiment analysis tools for software engineering research. Empirical Software Engineering, 2017, 22(5): 2543-2584.
- [31] Islam M R, Zibran M F. Leveraging automated sentiment analysis in software engineering//IEEE/ACM 14th International Conference on Mining Software Repositories (MSR), IEEE, 2017:203-214.
- [32] Ahmed T, Bosu A, Iqbal A. SentiCR: a customized sentiment analysis tool for code review interactions//Proceedings of the 32nd IEEE/ACM International Conference on Automated Software Engineering(ASE), IEEE Press, 2017: 106-111.
- [33] Ding J, Sun H, Wang X. Entity -level sentiment analysis of issue comments//Proceedings of the 3rd International Workshop on Emotion Awareness in Software Engineering(SEmotion), ACM, 2018:7-13.

- [34] Calefato F, Lanubile F, Maiorano F. Sentiment polarity detection for software development. *Empirical Software Engineering*, 2018, 23(3): 1352-1382.
- [35] Calefato F, Lanubile F, Novielli N. EmoTxt: a toolkit for emotion recognition from text//Seventh International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW), IEEE, 2017:79-80.
- [36] Novielli N, Girardi D, Lanubile F. A benchmark study on sentiment analysis for software engineering research//IEEE/ACM 15th International Conference on Mining Software Repositories (MSR), IEEE, 2018: 364-375.
- [37] Lin B, Zampetti F, Bavota G. Sentiment analysis for software engineering: how far can we go?//IEEE/ACM 40th International Conference on Software Engineering (ICSE), IEEE, 2018: 94-104.
- [38] Ortu M, Destefanis G, Adams B. The jira repository dataset: Understanding social aspects of software development//Proceedings of the 11th International Conference on Predictive Models and Data Analytics in Software Engineering (PROMISE), ACM, 2015, Article No. 1, DOI: 10.1145/2810146.2810147.
- [39] Russell J A. Core affect and the psychological construction of emotion. *Psychological Review*, 2003, 110(1): 145-172.
- [40] Destefanis G, Ortu M, Bowes D. On measuring affects of github issues' commenters//IEEE/ACM 3rd International Workshop on Emotion Awareness in Software Engineering (SEmotion), IEEE, 2018: 14-19.
- [41] Blaz C C A, Becker K. Sentiment analysis in tickets for it support//IEEE/ACM 13th Working Conference on Mining Software Repositories (MSR), IEEE, 2016:235-246.
- [42] Werder K. The evolution of emotional displays in open source software development teams: an individual growth curve analysis//Proceedings of the 3rd International Workshop on Emotion Awareness in Software Engineering(SEmotion), ACM, 2018: 1-6.
- [43] Ortu M, Destefanis G, Kassab M. Would you mind fixing this issue?//International Conference on Agile Software Development(XP), Springer, Cham, 2015: 129-140.
- [44] Graziotin D, Wang X, Abrahamsson P. Are happy developers more productive?//International Conference on Product Focused Software Process Improvement (PROFES), Springer, Berlin, Heidelberg, 2013:50-64.
- [45] Garcia D, Zanetti M S, Schweitzer F. The role of emotions in contributors activity: a case study on the Gentoo community//International Conference on Cloud and Green Computing (CGC), IEEE, 2013:410-417.
- [46] Calefato F, Lanubile F. Affective trust as a predictor of successful collaboration in distributed software projects//IEEE/ACM 1st International Workshop on Emotional Awareness in Software Engineering (SEmotion), IEEE, 2016: 3-5.
- [47] Khan I A, Brinkman W P, Hierons R M. Do moods affect programmers' debug performance? *Cognition, Technology & Work*, 2011, 13(4): 245-258.
- [48] Jurado F, Rodriguez P. Sentiment analysis in monitoring software development processes: an exploratory case study on GitHub's project issues. *Journal of Systems and Software*, 2015, 104(C): 82-89.
- [49] Novielli N, Calefato F, Lanubile F. The challenges of sentiment detection in the social programmer ecosystem//Proceedings of the 7th International Workshop on Social Software Engineering (SSE), ACM, 2015: 33-40.
- [50] Claes M, MntylM, Farooq U. On the use of emoticons in open source software development//Proceedings of the 12th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement(ESEM), ACM, 2018, Article No. 50, DOI: 10.1145/3239235.3267434.
- [51] Ferreira I, Stewart K. A longitudinal study on the maintainers' sentiment of a large scale open source ecosystem.