Soft Computing
https://doi.org/10.1007/s00500-022-06786-6

APPLICATION OF SOFT COMPUTING

Computational personality: a survey

Liang Yang[1] · Shuqun Li[1] · Xi Luo[1] · Bo Xu[1] · Yuanling Geng[1] · Zeyuan Zeng[1] · Fan Zhang[1] · Hongfei Lin[1]

Accepted: 6 January 2022

Abstract

Personality is a set of stable and tendentious behaviors, thoughts and emotions. How to measure personality more conveniently and accurately has always been a problem for scholars in related fields. With the rapid development of computer technology and the widespread popularity of social media in recent years, the research of computational personality has attracted wide attention of researchers in Computational Linguistics and psychology. Various methods, from statistical methods in psychology to machine learning and then to deep learning, have been proposed to deal with different areas of computational personality. In this paper, we first summarize the research framework of computational personality, and then review the current research progress of computational personality from the aspects of personality prediction, depression detection, suicide detection and happiness assessment, and provide the corresponding research resources for reference. Finally, we provide some possible research directions.

Keywords Computational personality · Personality prediction · Depression detection · Suicide detection · Happiness assessment

1 Introduction

Personality is the characteristic that draws the distinction between individuals. It plays a key role of guiding to our daily behavior and subjective perception. Our personality have a profound impact on our daily life. The study of personality is an important direction in psychology. The definition of personality varies with different scholars. Funder (1995) defined personality as a pattern of individual thinking, emotion and behavior, and the psychological mechanism which is hidden behind them. Adelstein et al. (2011) believed that personality describes persistent human behavioral responses to broad classes of environmental stimuli. Generally, personality is a collection of consistent psychological structure and behavioral characteristics for individual. Personality determines the pattern of individual behavior, and influences on the individual behavior, group behavior and even social development.

Psychology traditionally utilizes questionnaires and psychological tests for personality research, and conducts analysis by formulating psychological scale and collecting fragmented data of individuals. As such methods requires much manual labor, the volume of collected samples is limited. With the growing popularity of the Internet, especially the prevalence of social media, there are multiple channels for people to interact with each other, which also generates sufficient data for personality research. The data in the social media are now accessible with downloader like Spider. Individual personality can be reflected from the behaviors in the social media. The research of personality based on computing data is called computational personality.

Computational personality applies computing methods to personality analysis. The computing methods like data collection, feature extraction, machine learning enable the attributes of personality to be more accurately and effectively showcased. computational personality also introduces the psychological theory of personality to data analysis. This allows the application of personalized recommendations, early warnings of negative psychology, the detection of aggressive speech.

Our paper aims to comprehensively introduce computational personality and its latest progress. At first, we provide an overview of the research framework of computational personality, which summary a complete research process from resource to downstream application. Then, We focus on the four direction of computational personality: personality prediction, depression detection, suicide detection and

✉ Liang Yang
liang@dlut.edu.cn

1 School of Computer Science and Technology, Dalian University of Technology, Dalian, China

Published online: 13 February 2022

happiness assessment. These directions are the trend for computational personality, and have gained much attention. We sort out the relevant research of computational personality from the above four direction, and introduce the research progress, main challenges and the feasible research of computational personality in the future. After that, we review the data resources commonly used in computational personality research and related competitions. As computational personality is an emerging study with multiple interdisciplinary fields, there are still difficulties and challenges, which include data screening, privacy protection. We give a brief discussion on the ethical issue and the interpretability of the models on which is also worth working.

The research of computational personality has far-reaching social significance. Related applications are now presented at the detection of depression, early monitoring of suicide, identification of criminal suspects, and improvement of marital status. This work provides a comprehensive review and discussion of recent progress in computational personality, and gives some insights on the future direction. We hope that this can promote the development of computational personality.

The rest of this paper is organized as follows. Section 2 presents the related surveys on computational personality and its four subfield-personality prediction, depression detection, suicide detection and happiness assessment. In Sect. 3, we introduce the overall framework of computational personality research. In Sect. 4 , we comb and compare the existing works on the four subfield of computational personality in detail. In Sect. 5, we review the data resources commonly used in computational personality research. In Sect. 6, we discuss the ethical issue and the interpretablity of computational personality research. Section 7 is the conclusion of the paper.

2 Related works

There has been some survey research on computational personality, including the four aspects contained in this paper. Ilmini made a review (Ilmini and Fernando 2018) about computational personality in ICIIS 2017, and reviewed methods and theories involved in psychological traits assessment and evolution of computational psychological traits assessment with different machine learning algorithms and different feature sets. The review mainly focused on the different modalities: audio, video, pictures of faces, text and so on, basing on the Big Five theory. Zafar and Chitnis (2020) provided a detailed description and discussion on the definition of depression, depression rate, and causes of depression, and summarized and analyzed the data mining algorithms for depression content in social media. Gupta and Sharma (2021) aimed to identify the different machine learning algorithm methods, techniques, and approaches used by various studies related to depression detection on social media platforms by conducting a comprehensive review. Chiranjeevi et al. (2019) provided a comprehensive survey of human action recognition, machine learning techniques and various suicides prevention methods through which hanging attempts can be detected.

However, in this paper, we review the research of computational personality in the four aspects, which provides a different perspective from these works. We also provide an overview of the computational personality research framework. In addition, in this paper, we add the introduction and comparison of the technical methods used for computational personality in recent years.

Table 1 The Big Five Model

Personality traits	Feature
Openness to experience	Imagination, adventure, curiosity
Extraversion	Passion, vitality, dominance
Neuroticism	Anxiety, anger, impulsiveness
Conscientiousness	Rationality, responsibility, self-discipline
Agreeableness	Trust, honesty, obedience

3 Computational personality research framework

3.1 Theoretical basis of personality

3.1.1 The classification models for personality

The most common personality models for classification in psychology are the Big Five model (FFM or Big Five personality traits) and MBTI (Myers-Briggs Type Indicator).

The Big Five Model is one of the popular models for personality classification among researchers. It is now the five dimensions of personality characteristics that describe the high-level representation of individuals in psychology. These five dimensions or these five major personality traits are: Openness to Experience, Extraversion, Neuroticism, Conscientiousness, and Agreeableness. The features of each personality dimension are shown in Table 1. The Big Five Model describes the differences of individual personality, and also introduce a theoretical basis for the study combining psychology and computing. The model is of significance and dominance in psychology. The model utilizes numerical values to calculate personality types, which facilitates data processing through computers (Funder 2001).

Myers-Briggs Type Indicator, also referred to as MBTI, is another model of representative to study personality in

Table 2 MBTI scale

Dimension	Types	Abbreviated letter	Types	Abbreviated letter
Direction of attention	Extrovert	E	Introvert	I
Cognitive style	Intuition	N	Sensing	S
Judgment	Thinking	T	Feeling	F
Lifestyle	Judging	J	Perceiving	P

psychology. It is a scale-based personality evaluation model. Its theoretical prototype is the personality type which put forward by the founder of analytical psychology Carl G Jung (Pianesi 2013). The MBTI model divides the personality into four dimensions: extroverted and introverted (E/I), intuition and sensing (N/S), thinking and feeling (T/F), judging and perceiving (J/P), as shown in Table 2. The four dimensions are like four rulers. The personality of each individual will fall on a certain point of the ruler. Currently, there are some websites for the test of personality that can support MBTI[1].

However, the psychologists believe that the MBTI is too idealized (Capraro and Capraro 2002). As the results for the same person may vary with different time, MBTI is not in much of reliability. In contrast, the Big Five model is a better model for personality measurement.

3.1.2 The measurement of personality

Based on different personality theories, different models for personality measurement have emerged. Traditional personality measurement methods include self-report inventory such as the Minnesota Multiphase Personality Test (MMPI), the Cartel 16 Personality Factor Test (16PF), and AI Senk Personality Questionnaire (EPQ), Edward Personality Preference Scale (EPPS), California Psychological Questionnaire (CPI).

However, traditional personality measurement demands the cooperation of the measuring personnel and the person being measured. Otherwise, the accuracy of the measurement may not be fully guaranteed. With the popularity of social media, researchers have seen the feasibility of measuring the personality characteristics of users based on individual behavior in the social media.

Most of the data commonly used for personality analysis is text. Recently, researchers have begun to use multimodal technology to combine sound information, video information with text information to improve the effectiveness of personality analysis.

[1] https://www.16personalities.com/.

3.2 Research framework for computational personality

Based on personality theory in psychology, we also summarizes the research framework of computational personality. The research framework of computational personality is shown in Fig. 1.

At the resource level, computational personality refers to the data from the social media as sources. At the same time, computational personality draws on a variety of semantic resources in natural language processing to build the model of personality measurement.

At the theoretical layer, computational personality is guided by the Big Five Personality and other personality theories in psychology to realize psychology-oriented computational analysis and mining of personality.

At the algorithm layer, computational personality takes advantage of machine learning and deep learning to conduct personality analysis and prediction models.

At the application layer, the computational personality is applied to more fine-grained personality analysis research directions which include depression detection, suicide detection, and happiness assessment.

After integrating the four layers mentioned above, a platform of computational personality is established. In-depth analysis and application of users' personalities are carried out to achieve accurate and effective modeling.

4 Research on computational personality

With the increasing modern life pressure, we have to pay more attention to our mental health. However, as people may still have different discrimination and misunderstandings about mental illnesses, patients with mental illnesses in the early stage are afraid to seek medical treatment, and led to the deterioration of the disease. Computational personality are proposed to better understand people's mental health, achieve early detection and propose early treatment. In computational personality, the status of mental health has been studied in four direction: personality prediction, depression detection, suicide detection and happiness evaluation. The following content is the discussion of these four direction.

4.1 Personality prediction

Personality usually varies with different people. Different personalities can lead to different behaviors. The content users like to browse in the social media is also different. The portrait of personality can thereby be utilized to build the system of recommendation. In e-commerce business, we can recommend a suitable product to a person by analyzing his personality.

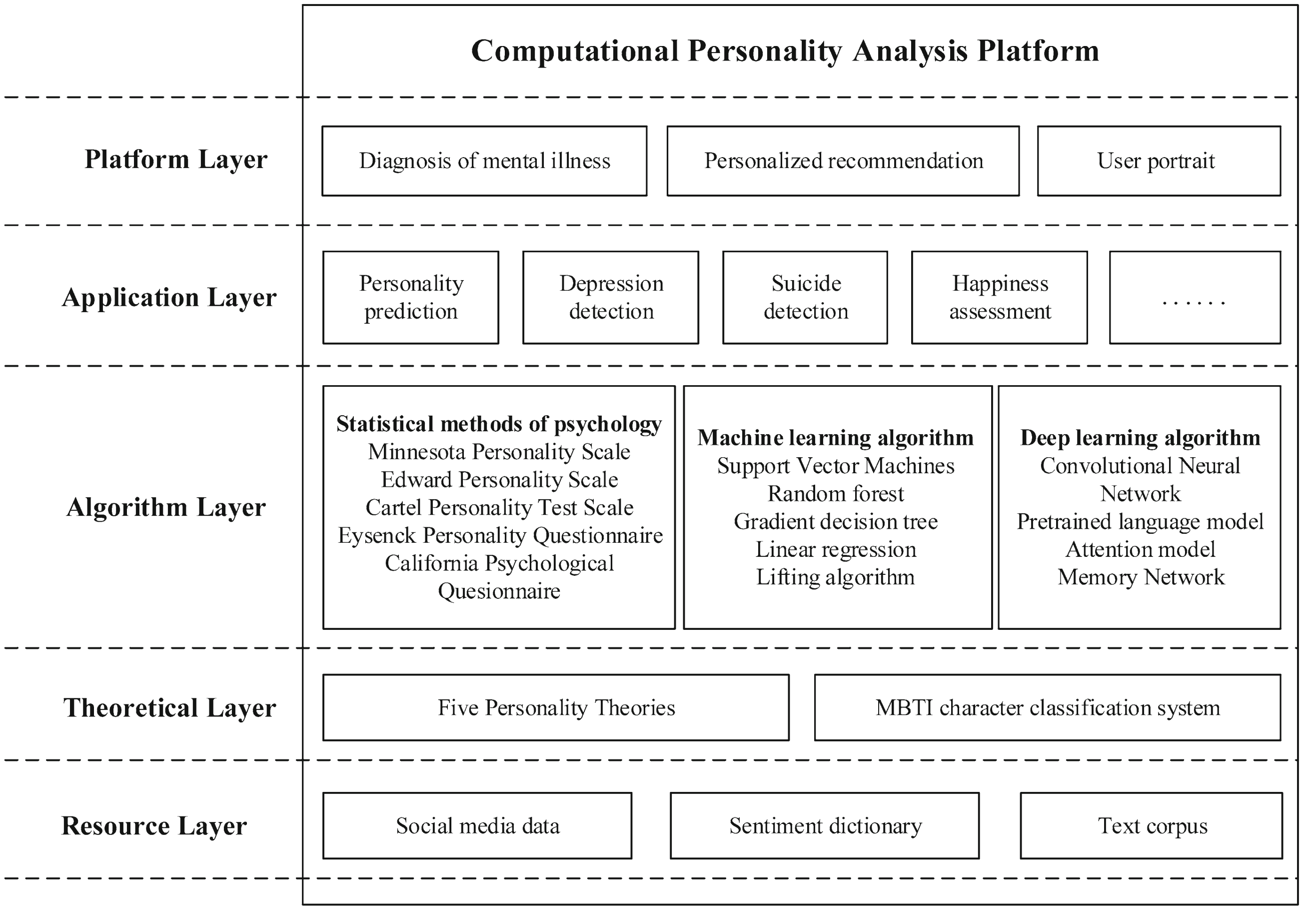

Fig. 1 Computational Personality research framework

4.1.1 Personality prediction based on big five model

The representative and early research of personality prediction based on the Big Five Model conducted by Shlomo et al. (2005). who used the support vector machine (SVM) and added lexical features to predict the "neurotic" and "extroverted" personality in the Big Five model. Argamon et al. Shlomo et al. (2005) used the paperwork of the students from the University of Texas at Austin between 1997 and 2003 as their data, and conducted a personality measurement on these students with Big Five model. Their research results showed that lexical features are applicable to the analysis of neurotic personality, but the accuracy of the prediction results of extroverted personality is not optimistic.

Besides paperwork, the computing data from social media for personality prediction is another source to conduct research. Zheng and Zuo (2016) referred to the data from the website myPersonality www.personality.org to conduct their research. A character prediction method combining information gain and semantic features is proposed to extract emotion words, parts of speech, tenses and other features from the text, select and weight the features, and map the text content to ontology concepts and calculate semantic relevance. Finally, based on lexical features and semantic features, support vector machines (SVM), K nearest neighbors (KNN), naive Bayes (NB) and other machine learning algorithms are used to predict personality. Compared with the general feature method based on LIWC(Linguistic Inquiry and Word Count) (Pennebaker and King 1999), their research produced a prediction with higher accuracy. Based on the Big Five Model theory, and with the data from the Chinese social media Weibo (much the same as Twitter), Liu et al. (2019) proposed an approach to carry out personality prediction with linear regression. Their accuracy of the methods reached 78.5%.

The studies mentioned above all refer to the text as experimental data, while other researchers believe that the interactive behaviors on social media (such as relay, likes.) can also be utilized as the source of data for personality prediction. Kosinski et al. (2013) believe that digital records such as Likes in Facebook can characterize personal attributes such as sexual orientation, race, religion, political orientation, personality characteristics, IQ, and happiness. This analysis is based on the data of more than 58,000 volunteers who provided their Likes data on Facebook, some personal data with detailed information, and their psychological tests. The model used dimensionality reduction to preprocess the data, and then input logistic regression/Linear regression to predict Likes' individual personality characteristics. The model correctly distinguishes 88% of gay and heterosexual men, and 85% of Democrat and Republicant supporters in the dataset. For the "openness" personality traits, the prediction accuracy is close to the accuracy of the standard personality test.

Jennifer et al. (2011) analyzed the information of users from Facebook to predict their personality. The information of users includes name, education level, marital status, density of social networks (that is intimacy with friends), and the number of words used to describe favorite activities, the number of joined groups, and political tendencies. Experimental results show that the Gaussian algorithm and M5 algorithm predict the user's RMSE personality score value is less than 0.13, which can accurately predict the user's personality.

There are also other researchers refer to images to analyze and predict personality. Fabio et al. (2014) analyzed the profile pictures of Facebook users. They believe that personality is related to the style of personal interaction. In their research, different machine learning algorithms such as SVM, naïve bayes, decision trees, logistic regression, nearest neighbors are utilized to test the effectiveness of image features in predicting personality and interactive style features. Effective machine learning algorithms can more accurately identify the personality characteristics of users through the avatar.

The research on personality prediction by multi-modal is increasingly extensive. Marcin et al. (2016) believe that in many personalized retrieval and recommendation systems, the characteristics of users' personality are also significant. They thereby integrated text, image and users' meta features related to their reputation and influence, and analyzed two different SNS (Twitter and Instagram networks) at the same time. Their prediction error of each personality trait was uniformly reduced.

Onno et al. (2018) used multi-modal to predict personality based on the Big Five Personality with voice, text, and video (mainly facial features). Referring to the acoustic and prosodic information, the original sound waves are input into the convolutional neural network (CNN); for the data of video, the picture of one frame is randomly selected from each video. The feature of personal appearance in this frame is used to identify the personality with the VGG-FACE CNN model; for text information, using word2vec to convert them into word vector and input them to the CNN model. Their results show that the multi-modal fusion improves the performance by 9.4% compared with the mono-modal with video feature.

Aslan et al. [15] proposed a multi-modal method based on deep learning, extracting the features includes speaking style, facial expressions, body movements, language from the video, and use pre-trained deep convolutional network (Resnet and VGGish) to extract high-level features, and also LSTM to gather time information. Their average accuracy of the five features in the Big Five Model was the best result in the year of their research.

Lynn et al. (2020) proposed a neural model which used message-level attention to recover highsignal messages from noisy data for personality prediction. And they demonstrate that models with message-level attention outperform those with word-level attention.

Christian et al. (2021) proposed a multi model deep learning model architecture with a pre-trained language model BERT, RoBERTa, and XLNet to extract features for personality prediction system. They also combined multiple sources of social media data to increase the number of datasets for better performance. Their method produced better performance than previous studies that predicted personality traits.

There are increasing numbers of researchers that conduct personality prediction research based on the Big Five Model. Common methods for personality prediction are extracting features and classifying with machine learning algorithms, and using natural language processing to embed text words into neural networks. At the same time, the pre-trained language model (such as Bert) has also brought new vitality to computational personality. In addition to text information, multi-modal combined with text, sound, image to predict personality.

4.1.2 Character prediction based on MBTI personality scale

Myers-Briggs Type Indicator (MBTI) is popular among the non-scientific communities. Many online users refer to MBTI personality scale to analyze their own personalities and share their discussions on social media, which produce self-assessment data about MBTI on social media such as Twitter.

One of the works of representative with MBTI include (Ben et al. 2016) who created a new MBTI-based corpus from Twitter, which is applicable with six languages. They referred to linear SVC in sklearn and logistic regression to give a baseline, and their results are also effective in gender recognition.

Kosuke et al. (2019) believe that textual information on social media and individual behaviors (such as likes and reposts) can support the personality prediction. They find that many users form social media often browse the posts of other users. However, these users rarely publish their own posts. And these users implicitly express their opinions and preferences through the behavior such as like and share. Therefore, Kosuke et al. crawled the data from Twitter to build a dataset, based on MBTI personality model indicators, and with BOW, SVD, DBOW, Co-occurrence to analyze personality.

Table 3 Personality prediction based on Big Five model

Study	Input	Architecture	Datasets
Shlomo et al. (2005)	Text	SMO learning algorithm	Student essay
Kosinski et al. (2013)	Facebook likes	Logistic or Linear Regression	–
Jennifer et al. (2011)	Text	Gaussian and M5 algorithm	Collect from facebook
Fabio et al. (2014)	Image	BoVW+ML	Collect from facebook
Marcin et al. (2016)	Image and text	–	Twitter and Instagram
Onno et al. (2018)	Audio, text and video	Multimodal fusion	ChaLearn First Impressions
Süleyman and Uğur (2019)	Video and text	Multimodal fusion	ChaLearn First Impressions V2
Ilmini and Fernando (2018)	Text	Sequential model with attention	Facebook
Lynn et al. (2020)	Text	PLMs	MyPersonality, twitter data

Online education has become an emerging tech of the modern education system. The effects of online education vary with the students with different personalities. Mohamed et al. (2015) used a number of 10 classification algorithms based on MBTI theory to analyze the dataset collected from the German Business School in Cairo, Egypt to help students realize their personality characteristics, so that their study habits become more effective.

Matej and Jan (2018) referred to Reddit to construct the dataset with MBTI - MBTI9k, which uses personality extraction and three machine learning algorithms (SVM, LR, MLP) to classify personality. The F1 value in the "intuition and sensing" (N/S) personality dimension reached 82%, but only 67% in the "thinking and feeling" (T/F) personality dimension, so it may be necessary to use a deep learning model to improve the performance index in T/F character dimension.

In recent years, e-commerce has developed rapidly, so it is particularly important to provide users with more accurate recommendation services. Chen et al. (2019) proposed a user model based on consumer personality. According to different consumer personality, users are divided into five types: compliance, conservative, free, eccentric and economic. Their research selected 500 actual commodity resources in the category of clothing from an e-commerce website, and applied keyword-based method of representation, traditional text-based method of representation, and ontology-based representation of consumer personality method for the recommendation results. Their results show that ontology-based representation method is more accurate than the former two, and the recommended products through this method is more in line with user needs.

4.1.3 Insights

Most of existing personality prediction is based on the open-source data from the social media, and based on the Big Five model and MBTI. Recent researchers combine personality theory with social media to study computational personality. Through the portrait of personality, we are able to better understand the users' behaviors of social media. Through personality information, we are able to better understand the users' behaviors of social media. Computational Personality can be used to do personalized recommendations, public opinion analysis, and user psychological warnings (such as suicide detection, depression detection, happiness evaluation, etc.).

4.2 Depression detection

Depression is one of the prevalent diseases across the world. The detection of depression in the early stage can effectively control the disease without the deterioration of the disease. However, many patients are unwilling to consult a psychologist because of their mental burden. Many researches now turn to computational personality to detect the users with potential depression.

Depression detection is of significance in computational personality. It refers to obtaining user-generated information (UGC), with natural language processing to detect and predict the depression of users. Depression detection takes UGC as input, while the output is the probability of whether the user has depression.

4.2.1 Depression detection based on offline questionnaire

In psychology, as early as 2006, Cloninger et al. (2006) studied the role of personality for individuals suffering from depression; in sociolinguistics, Oxman et al. (1982) pointed out that patients can be divided into depression and paranoia with their record in social media. Accordingly, their research on depression detection is only based on the written text.

The analysis of the written text through LIWC also revealed some clues about the prediction of psychosis and mental illness (Rude et al. 2004). Resnik et al. (2013) used the paper materials collected by Rude et al. (2004). These paper materials were written by students with the title of "the

Table 4 Character prediction based on MBTI personalityscale

Study	Input	Architecture	Datasets	Performance
Shlomo et al. (2005)	Text	LinearSVC	TWISTY	Average F1 score 59.61 on ES language
Kosuke et al. (2019)	Text	LinearSVM	Tweets in janpanese	Average AUC score 0.6893
Mohamed et al. (2015)	Text	OneR, Random forest	LMS	Recall score 0.974, 0.933
Matej et al. (2018)	Text	LR_all, MLP_all	MBTI9k	Average F1 score 75.15, 75.1

most impressive thoughts and feelings in college life", combined with the Beck Depression Scale (Beck et al. 1961) to measure the severity of depressive status. They also utilized the LR model to extract the features of LIWC and LDA (Blei et al. 2003), and to make the combination of the two features. Finally, they predict whether the students writing these paper materials suffers from depression. Their research referred to computational personality methods to detect textual materials for the first time, which gave the possibility of using social media as a carrier for depression detection.

The above research provides some ideas for the detection of depression and their summaries are given by Table 5, but they also have some limitations. The size of samples are limited. Samples have similar characteristics. These samples thereby may not represent the majority of people with depression. In addition, these studies are based on questionnaires and similar surveys, relying on retrospective self-reports of depression patients' and health observations. Asking for much labors, such surveys usually last for months or even years. All in all, it has the disadvantages of wasting resources and being limited by time granularity, which makes it impossible to detect depression effectively in time.

In terms of depression detection, traditional machine learning for depression detection research has made significant progress. Similarly, the methods of feature extraction have become diversified. From studying the extraction of text features in the early stage, to adding information such as time dimension and emotional dimension later, feature extraction from different angles has different contributions. Deep learning models are also used in depression detection, but there is still a lack of breakthrough work. On the other hand, most of the existing researches are aimed at studying English speakers, and the detection of depression in other languages is still at the preliminary stage, so there is still much room for improvement in cross-language depression detection. With the diversification of social media forms, a small number of research works have taken a different approach, using user-generated text information, additional picture information and sound information as a dataset for depression detection, and performing depression detection from a multi-modal perspective, which is more clinical reliability.

4.2.2 Depression detection based on social media

• Depression detection combined with written text

Moreno et al. (2011) proved that status updates on Facebook can be used to reveal the symptoms of severe depressive episodes, while (Park et al. 2012) found the evidence on the Twitter that people with depression tend to post their feelings of depression and depression treatment on social media. Therefore, we can use the data from social media related to users' psychological activities and social environmental activities to detect the possibility of users having depression in a time-grained manner; However, due to the lack of relevant social media datasets, and the fact that users cannot be judged to suffer from depression based solely on social media data, so the datasets used by most research institutes at this time are constructed by combining social media dynamics with face-to-face interviews or questionnaire surveys.

De et al. (2013) explored the potential for depression detection on social media by collecting data on Twitter. They established the model by combining the behavior of users with depression and the behavior of users without depression. (The users with depression show reduced social activities, increased negative emotions, high self-concern, and increased expression of religious thoughts.) However, this study is also lack of annotated datasets, and the SVM classifier is only with 74% accuracy. Park et al. (2013) also used the data from Twitter to detect whether users are suffering from depression. They conducted semi-structured face-to-face interviews with 14 active users on Twitter, and had a qualitative analysis to study the behavior and perception differences between depressive users and non-depressive users in social media. The study concluded that depressive users prone to regard social media as a tool for emotional sharing, while non-depressive users use social media as a platform for only sharing information.

The above research work combines social media dynamics, questionnaire information, interview information, and actual psychological activities to obtain results. However, the collection of interviews and other information requires man labors and material resources, which restricts its development to a certain extent.

• Depression detection using only social media information

Table 5 Research on depression detection based on offline questionnaires

Study	Input	Architecture	Datasets	Performance
Cloninger et al. (2006)	Text	Questionnaire statistics	TCI and CES-D completed by 631 adults	TCI sorces:0.75 to 0.85
Rude et al. (2004)	Text	Questionnaire statistics	Student essays	BDI:20.05;4.17;3.58 IDD-L:25.60;40.59;1.18
Resnik et al. (2013)	Text	LIWC,LDA	Student essays	r= 0.459

Inspired from the study by Malmasi et al. (2016), and based on their study of sentence structure of "I am/I used to be/depression", the texts can be marked as a depression-oriented article or a standard article. Their study directed the goal to obtain the annotated data from social media. After their approach was proposed, most of the research turned to focused on analyzing the text from social media. Many of the features are commonly used to understand the mental state of individuals: LIWC are commonly used to extract the mental features of individuals. The work of Wang et al. (2017) extracts the features of first, second, and third person pronouns, words related to the perception process, or positive and negative emotional words. When conducting sentiment analysis on text, tools as OpinionFinder (Wilson et al. 2005), SentiStrength (Thelwall et al. 2010), and ANEW (Bradley and Lang 1999) are used to quantify perceptual and emotional attributes of text in expression. In addition, emojis and images are also used to detect positive and negative emotions of the posts in social media (Kang et al. 2016). As part of content analysis, in order to extract topics from user-generated content, various types of topic modeling are integrated. One of the popular models is LDA.

When making predictions on datasets, most of the models are trained by supervised machine learning models. Shen et al. (2017) constructed fully labeled depression and non-depression datasets, and extracted six feature groups related to depression. These features cover the standards of clinical depression and the behaviors on social media. With these feature groups, they propose multi-modal dictionary of depression. And their effectiveness of the method is proved to realize the depression detection for the users on Twitter. Hiraga (2017) explored the relations between the text features and the ones with depression by supervised machine learning. In his research, Japanese blogs with general themes were referred to as datasets to extract text features. For the classification of blogs, he used Character n-grams and Token n-gram, the multinomial naive bayes (NB) classifier presented by scikit-learn (Pedregosa et al. 2011), linear support vector machines (SVM) classifier and logistic regression (LR) classifier; Hiraga's research shows that classification tasks at the author level can achieve an accuracy of 86.4%. For text-level classification tasks, the highest accuracy rate is 75.5%; but because of the small scale of data, it is unknown whether it can still be effective on large-scale datasets.

Different from traditional machine learning methods, Yates et al. (2017) used a neural network model to detect depression and described the close relationship between self-harm and depression. According to the self-reported depression diagnosis (RSDD) of Reddit, a dataset for identification was constructed. Unlike the previous extraction of specific features and the use of supervised machine learning models, their research started from the text content and proposes an improved neural network CNN (Kalchbrenner et al. 2014) model, which achieves excellent performance on F1. However, due to the anonymous environment of Reddit for users, it is impossible to ascertain whether the data is authentic. In addition, the study also ignores the users who have not self-report, and ignores the hidden tweets sent by the users with depression.

After a series of research, the basic methods of research for depression detection on social media has come to a standstill. Most of the basic research are collected data and model prediction within a short time period, ignoring the function of discrete emotions. Chen et al. (2018) used emotional characteristics and time series method to detect depression on social media users. They used Twitter posts as a dataset, extracted the features of eight basic emotions. Then, the time series measurement method is used to analyze the emotion measurement results, and a set of time feature is obtained. The depression is detected by the combination of emotion features and time features. According to their experiment, the accuracy rate of the prediction results extracted with only emotion features is 87.27%, which is higher than the results of previous studies. In addition, combined with time series features, the accuracy rate can reach 89.77%. Their experiments have shown that emotional features can reveal an individual's mental state changed over time. The changes of the mood of individuals can also provide more information for the detection of depression. Although the annotation of psychological data requires professional training, their approach brings a new perspective for depression detection.

Zogan et al. (2021) believes that using all information about user behavior and their linguistic patterns including users' social interactions to predict depression may harm the overall efficiency and effectiveness of the model because the contents contain irrelevant information to depression. Therefore,they proposed a novel framework for automatic depression detection that initially selects relevant content through a hybrid extractive and abstractive summarization strategy on the sequence of all user tweets leading to more fine-grained and relevant content. The content then went to their novel deep learning framework leading to better empirical performance than existing strong baselines.

Unlike previous studies only with textual materials as input, Samareh et al. (2018) proposed a method for the prediction of the severity of depression through multi-modal method. With the dataset of AVEC 2017 (Ringeval et al. 2017), their approach showed a well-performed result. The AVEC 2017 dataset is composed of audio, video and text information. By extracting the features of audio, video and text information separately, the random forest classification model is used to obtain the score of each feature. The fusion of features for the decision model based on the confidence is used to obtain the final forecast results. Similarly, Gui et al. (2019) chose to study from text information and visual information, using reinforcement learning methods to screen related tweets, and effectively merged the features of text and video. An et al. (2020) found that the global topic information obtained from user-generated content (such as text and images) is essential to improve the performance of depression detection tasks. Therefore, they proposed a modality-agnostic topic model that can learn from discrete text Mining topic clues from signals or continuous visual signals as an auxiliary task to improve the accuracy of depression prediction.

In addition, depression detection has been proven effective on platforms such as Twitter and Facebook. However, due to cultural differences, it is not possible to directly apply some research methods to social media in other languages, such as Weibo which uses Chinese as the basic language. It may not work well due to the lack of annotated datasets. Shen et al. (2018) proposed a cross-domain deep neural network model with feature adaptive transformation, using the data from Twitter data as sources. They studied the detection of depression in a specific target domain (such as Weibo). This method is effective and can detect depression in cross-domain languages. Although good results can be achieved on some datasets, they still lack of the analysis of combination with the online and offline situation. Therefore, it cannot be directly applied to clinical diagnosis. The summary of the above research is shown in Table 6.

4.2.3 Insights

Since its development, the detection of depression on social media has achieved considerable progress. With the increase of the accuracy rate, it can be used to predict depression so that patients may receive early treatment.

For the future research directions, the following several future research directions are proposed: Researchers can use a multi-dimensional methods for depression detection to be more practical in actual situations; with the continuous improvement of the accuracy of depression detection, how to apply depression detection to clinical detection has also become a major research topic in the future, such as through the information created by the user, combined with the clinical symptoms of depression, to determine the severity of depression, and timely intervention.

4.3 Suicide detection

According to the latest data from the US Centers for Disease Control and Prevention (CDC), suicide is the second leading cause of death between the ages of 10-34 and the fourth leading cause of death between the ages of 35-64. This rate of suicide is still on the rise. According to statistics, 287,000 people die from suicide each year in China, and 2 million people attempt suicide. Suicide is a serious public health problem, which can bring about direct and indirect economic, social losses. Traditional suicide risk assessment research mainly uses psychological tests, questionnaires and other analysis methods, but in actual cases, these methods still have certain limitations. As social networking platforms have given people more opportunities to reveal their feelings and opinions in the virtual community in recent years, then through social networks, they can actively find individuals with potential suicidal tendencies and analyze them to provide early warning. At present, social media-based suicide risk assessment studies often use four-label categories(no risk, low risk, moderate risk, and high risk) to classify suicide users, and predict their suicide risk through their expression and behavior information on social media.

4.3.1 Suicide detection based on questionnaire

Psychology researchers have developed some psychological measures to obtain suicide risk: Bagge and Osman (1998) suicide probability scale, King-wa et al. (2007) adult suicidal ideation questionnaire, Keith et al. (2015) suicide effects behavioral cognitive scales. Each of the scale mentioned above has its own applicable group. These psychological scales have shown decent results in actual use, and can be referred in the suicide detection. Sueki (2015) studied on the correlation between suicide-related Twitter and suicidal behavior. Participants answered a self-managed online questionnaire that included the questions about the usage of Twitter, suicidal behavior, depression and anxiety. The survey results showed that the analysis on text of Twitter is able to identify young Internet users with suicidal possibilities.

However, suicide detection based on questionnaires has its own limitations. The scales require respondents to fill

Table 6 Research on depression detection based on social media

Study	Input	Architecture	Datasets	Performance
Moreno et al. (2011)	Text	Demographics characteristics	Facebook profiles	MDE:2.5%
Jashinsky et al. (2014)	Text	SVM, LIWC	Twitter	ACC:74%
Coppersmith et al. (2016)	Text	SVM, RF, RBF Lexical Features Syntactic Features	2016CLPsych SharedTask	F:87% ACC:91%
Kang et al. (2016)	Text and Image	SVM, WordNet VSO	Tweets	F1:86.72% ACC:90.04%
Shen et al. (2017)	Text and Image	LIWC, VAD LDA MDL	Depression, Non-Depression, Depression-candidate Datasets	F1:85% ACC:85%
Hiraga (2017)	Text	NB, SVM, LR	Japanese Blog posts	ACC:95.5%
Yates et al. (2017)	Text	CNN	RSDD	F1:51%
Chen et al. (2018)	Text	EMOTIVE, LIWC SVM, RF, LR, NB Mean Momentum, Mean Differencing	Twitter	F1:91.8% ACC:93.6%
Gupta and Sharma (2021)	Text	BERT, K-meas BART, BiGRU, CNN Attention	Depression, Non-Depression, Depression-candidate Datasets	F1:91.2% ACC:90.1%
Samareh et al. (2018)	Audio, Video, and Text	COVAREP, AFINN RF	AVEC2017 dataset	RMSE:4.81
Gui et al. (2019)	Text and Image	COMMA, GRU,V GG-Net, RL	Textual Depression, Multimodal Depression Datasets	F1:90% ACC:90%
Zogan et al. (2021)	Text and Image	BERT, VGG UGC Topic model	Textual Depression, Multimodal Depression Datasets	F1:84.2% ACC:84.2%
Shen et al. (2018)	Text and Image	DNN, DFC, FNA	Weibo Dataset DT Twitter Dataset DS	F1:78.5%

out evaluation forms or participate in interviews, but these scales may only target on certain factors that affect suicide, or only on certain groups. Suicides who rarely seek professional help may not be recognized. At the same time, it is time-consuming and labor-intensive, making it difficult to perform large-scale real-time suicide detection tasks. It still needs further research to improve the accuracy and efficiency of suicide detection. The introduction of computer science with larger scale of data and computing power is able to realize a universal and practical detection method.

4.3.2 Suicide detection based on the data from official institutions

Unlike the psychological questionnaire, the data from official institutions is generally in large scale and has many dimensions. However, such data also contains many unrelated dimensions for suicide analysis, so it is more relied on computer technology in data processing. Colin et al. (2017) used an anonymous electronic health record (EHR) dataset with a size of 500,000. They also referred to the random forest (RF) to perform a binary classification of suicide risk. The F1 score reached 86% and the recall rate reached 95%. Harish et al. (2017) used a deep neural network to predict suicidal thoughts. They also referred to the anonymous HER dataset to predict suicide among teenagers. Their model obtained 70% true positives and 98.2% true negative. Payam et al. (2016) used the suicide dataset in Iran and utilized SVM, LR, artificial neural network (ANN) and other traditional machine learning methods to assess suicide risk. They found that gender, age, jobs are important factors affecting the idea of suicide.

Suicide detection based on official statistics can explore more dimensions that affect suicide thoughts. However, the construction of such dataset requires much labor. And it is difficult to apply to the real-time suicide detection. Some other research (John et al. 2010; Jones and Bennell 2007) applied nature language processing algorithms on the suicide notes, which also faced limitation of applicability. With the widespread of social media, a large number of unsupervised corpora and social network features further provide comprehensive information.

4.3.3 Suicide detection based on the data from social media

The datasets based on social media research mainly come from community platforms such as Reddit, Facebook, and Weibo. Recent research also focus on identifying the authenticity of suicide information and detecting suicide information on social media. The former mainly analyzes how to determine the authenticity of some suicide-related datasets, such as the speech in the suicide sub-community of Reddit; the latter focuses on the identification in daily tweets (that is, daily published text) posted by users to recognize texts and users of high suicide risk.

Tim et al. (2013) applied text sentiment analysis to both posts and the comments in Chinese online forums, and to identify suicide texts. Xiaolei et al. (2015) based on the data form Weibo (a Chinese social media), with word embedding and psychological standards, using topic models to identify suicide thoughts, and constructed a prototype system for suicide detection to monitor suicide texts in real time. Burnap et al. (2017) performed suicide detection classification to distinguish suicide-related topics, such as whether it is a real suicide or just simply mentioning it. Lei et al. (2019) studied the application of tree holes in Weibo to conduct suicide risk detection, using a two-layer attention mechanism to capture the change of suicide thoughts from individual blog. Based on the word embedding and attention mechanism, a more effective suicide risk detection model is proposed which outperforms their well-designed benchmark, with an accuracy rate of 91%.

The feature extraction of text and other forms of data have also gained the focus by the computational personality researchers. (Jashinsky et al. 2014) used SVM to predict the suicide possibilities of individual in a certain timeline, with the term frequency-inverse document frequency (TF-IDF) of each tweet, and with the input feature of each tweet's word frequency, unique word count, average word count. Munmun et al. (2016) extracted the features from language, vocabulary, and network, using these features to describe the characteristics of patients with mental health, and predicting their risk of conducting suicide. From analyzing the content of self-reported posts on Reddit, they obtained the mental health status of relevant users, and used matching mechanism of propensity score, to predict the possibilities that users will share suicidal thoughts in the future. Lipeng and Wenai (2019) extracted dictionary features and language features based on the data from Weibo, and verified that the performance of the model based on n-gram features was improved in a certain extent. Sawhney et al. (2020) proposed a time-aware transformer which focused on identifying suicidal intent in English tweets by augmenting linguistic models with historical context.

Recently, some researchers paid attention to the emotional polarity of the target text, which has gained significant improve on suicide detection. Sawhney et al. (2021) proposed a time-and phase-aware framework that adaptively learns features from a user's historical emotional spectrum on Twitter for preliminary screening of suicidal risk. In addition, they Sawhney et al. (2021) jointly leveraged a user's emotional history and social information from a user's neighborhood in a network to contextualize the interpretation of the latest tweet of a user on Twitter. Li et al. (2021) presented an approach to detect suicide ideation based on multi-task learning which introduced emotion information.

Some researchers also construct new datasets based on the original datasets by introducing domain knowledge and adding data features. Gaur et al. (2019) introduced medical knowledge to predict the suicide risk of individual, and suicide ontology to develop a suicide risk severity dictionary, and also created a Twitter dataset. They expanded from the four dimensions to five dimensions for classification. Rohan et al. (2019) discussed the use of the information of user behavior on social media by inputting text features and embedding the features of social relation graphs. And they developed a manual-annotated dataset in Twitter. Their results verify the effectiveness of the proposed SNAP-BATNET model in suicide detection.

Besides the classification and recognition of suicide texts, some researchers have also tried to expand the scope of suicide detection, such as exploring the causes of suicide. Jingcheng et al. (2018) used deep learning methods to detect the source of mental stress that led to suicide, and used a convolutional neural network (CNN) to construct a binary classifier to identify suicide tweets. Once suicidal tweets are detected, they use recurrent neural networks (RNN) for named entity recognition (NER) to mark the source of mental stress in tweets classified as suicide, while achieving text recognition and causal reasoning.

In summary, the suicide detection research on social media has achieved certain results, and due to the open access of data and the development of text mining algorithm, the research prospects are still broad.

4.3.4 Insights

The current suicide detection research mainly focuses on using data such as texts posted by users on social media to classify users' suicide risk, and further includes the identification of suicide triggers and the judgment of the authenticity of suicide notes. The method level is generally to extract features of text, graphs and other data and classify them with machine learning models. With the rapid development of deep learning, various end-to-end suicide risk identifica-

Table 7 Research on suicide detection based on social media

Study	Input	Architecture	Datasets	Result(F1)
Tim et al. (2013)	Text and graph	Graph-based algorithm	Web forums	–
Xiaolei et al. (2015)	Text	Topic model	Weibo	0.80
Burnap et al. (2017)	Text	Rotation Forest algorithm and Maximum Probability vote	Twitter	0.728
Lei et al. (2019)	Text	Pre-trained word embedding and Attention	Tree hole on Weibo	0.91
Jashinsky et al. (2014)	Text	Keywords and phrases	Twitter	–
Munmun et al. (2016)	Text	Linguistic and network features	Subreddit	–
Lipeng and Wenai (2019)	Text	Random Forest	Tree hole on Weibo	0.93
Gaur et al. (2019)	Text	CNN	Reddit	0.65
Rohan et al. (2019)	Text and graph	BiLSTM with Attention	Twitter	0.65
Jingcheng et al. (2018)	Text	LSTM+CRF and CNN	Twitter	0.83
Sawhney et al. (2020)	Text	Time-aware transformer	Twitter	0.80
Sawhney et al. (2021)	Text	Phase-aware EmoBERT	Twitter	0.81
Sawhney et al. (2021)	Text and graph	Hyperbolic graph convolution network	Twitter	0.79
Li et al. (2021)	Text	BERT-BiLSTM with Adversarial and multi-task learning	SID	0.96

tion models and some online suicide risk detection models have emerged. It is foreseeable that as the research on suicide detection continues to deepen, we can more accurately identify people with suicidal tendencies and conduct timely psychological counseling to prevent suicidal behavior.

Furthermore, data sources and forms of suicide detection are constantly diversifying. Social graph data has attracted the attention of researchers. Suicidal thoughts are easy to spread in social networks, Therefore, studying the communication method and process of suicidal ideation can effectively improve the accuracy of suicide detection, and even curb the spread of suicide beliefs. A very common cause and precursor of suicide is the appearance of depression. It can be combined with depression detection to study the relationship with depression and suicide, which is helpful to improve the accuracy of suicide detection.

4.4 Happiness assessment

As society pays increasing attention to people's mental health, happiness has gradually become a popular research direction. computational personality studies the research on happiness by using natural language processing and other techniques to mine the information related to happiness in the text, such as the source of happiness, strength of happiness, assessment of happiness, etc. Happiness assessment aims to measure a person's happiness and quantify the strength of happiness.

Happiness mainly refers to the overall emotional and cognitive evaluations measured by people on their quality of life. It has three characteristics: subjectivity, stability and integrity (Diener et al. 1984). Diener et al. (1984) proposed that the sense of happiness consists of emotional dimension and cognitive dimension, that is, having more positive emotions but less negative emotions, as well as satisfaction with life. This idea reached a consensus between psychologists on the meaning, dimensions and measurement methods of happiness.

In the study of happiness, personality is one of the indicators for predicting happiness (DieNer et al. 1999). Based on interaction model proposed by Diener, it is believed that personality affects people's behavior and attitudes. This is also a theoretical basis for computational personality in psychology.

The research on the evaluation of happiness is mainly based on the scale in the early stage, and is turned to the text in the later period.

4.4.1 Early research based on scale

The study of happiness generally began in the United States in the 1950s. From the perspective of its development back-

Table 8 Research of happiness assessment

Author	Data sources	Method	External resources
Sukjin et al. (2013)	Flickr	Dictionary matching	The labMT 1.0 dictionary of happiness index Dodds et al. (2011)
Hao et al. (2014)	Weibo	Scientific indicators PANAS & PWBS	–
Qi et al. (2015)	Blog	Scientific indicators PANAS	UGC of online grassroots blogs
Shrey et al. (2017)	Reddit	Build indicators Mental Happiness Index (MWI)$^+$	–
Singh et al. (2017)	Twitter	Build indicators Net Sentiment$^+$	SentiWordNet

ground, one is due to the continuous improvement of people's quality of life, and the other is the rising concern of psychological health. Since the 1970s, researchers have expanded the research on this subject from the philosophical level to the scientific level. In this process, the explanation theory of happiness directly influenced people's research direction (Yan and Jun 2004). Based on the theory, research has turned to quantify happiness and explore methods to improve people's happiness.

The tool in the early stage for measuring happiness was the Affect Balance Scale (ABS) compiled by Bradburn in 1963. The scale has a total of 10 topics, including two dimensions of positive and negative emotions. Based on the total scores of these two dimensions, the main test speculates on the subject's recent emotional state, and further speculates on their subjective happiness level. Some researchers have suggested that subjective happiness is an individual's long-term stable rather than temporary emotional state. ABS evaluates the individual's temporary emotional state, so he doubts the use of ABS results to predict subjective happiness (Xingyu 2017).

After this, many happiness measurement tools have appeared, such as the DT scale (Delighted-Terrible Scale), the Memorial University of Newfoundland Scale of Happiness (MUNSH), and the life satisfaction scale (the Satisfaction With Life Scale, SWLS), etc. Lei et al. (2019) analyzed the structural dimensions of different existing happiness measurement tools in their paper "Evaluation and Prospect of Happiness Measurement Index System", and found that only 24% existing theories were used to develop measurement tools. The construction of theory-oriented index system can often make the test results more objective, fair and accurate. The analysis of psychological measurement features shows that the measurement of reliability and validity in each dimension of happiness are different. However, less than half of the measurement tools have a good reliability through retest.

Since the study of happiness has just started, people in this period mainly used scales to measure and evaluate happiness. Its effectiveness is closely related to professional psychological theory. However, the scale is dependent on the user, and the collection of statistical work is time-consuming.

4.4.2 Recent research based on text

In the era of rapid development of computers, questionnaires and social media are gradually being used to measure and evaluate happiness, due to the evaluation of happiness are closely related to the individual characteristics. The questionnaire survey is not only directly using the happiness measurement tools mentioned above, but also having the data containing the descriptions related to happiness. The public happiness dataset HappyDB[2] is released by (Asai et al. 2018) in 2018. Its aim is to conduct a questionnaire survey on workers in a certain factory, and record them in a certain period of time. A total of 100,000 happy moments were collected. Social media collects text data related to happiness from social media platforms such as Twitter, Flickr, and Blog. Due to the privacy protection of social media for users, datasets collected from social media are generally built by themselves and will not open to the public. However, there are still a few public datasets, such as Ren-CECps-SWB 2.0[3], a Chinese dataset published by Qi et al. (2015) and collected on China Grassroots Blog.

In terms of methods, researchers often use emotional dictionaries to assist evaluation. Table 8 summarizes the related work in recent years:

From the perspective of data sources, the datasets mentioned are all built based on social media. On the one hand, social media generates data in real time by users, which is directly related to users. It is easy to obtain, which is helpful for researchers to analyze. On the other hand, social media platforms pay attention to the privacy of users, and the collected data is not easy to open to the public, so you need to build the dataset in your own hand.

All the works listed above is based on sentiment analysis. As happiness is also a positive emotion, you can use some emotion dictionaries to classify emotions. Additional

2 http://rebrand.ly/happydb.

3 https://github.com/KGBUSH/Ren_CECps-Dictionary.

information from the data source, such as the tags will also provide to the user to express emotions in the blog, which can be used for the analysis by the researchers.

Happiness is a relatively complex emotion. As for the final evaluated indicators, it can be seen that there have been some changes over time. In the beginning, it was just a simple match with the emotion dictionary to have a rough estimation. Simple dictionary matching is not enough to describe happiness in depth. Later, using some scientific indicator methods, such as PANAS, PWBS, etc., and finally building a rating indicator for happiness assessment, it provides better methods for the evaluation of happiness.

4.4.3 Insights

The research on the evaluation of happiness starts with the scale. However, the scale requests the time-consuming design of professionals, and the data collection process is also cumbersome. Its efficiency can be largely improved by the computer science.

In the future, the study of happiness will be more in-depth and detailed. Here are two prospects for the assessment of happiness: 1. Data sources will tend to diversify, not only questionnaires and social media, but also richer data such as wearable devices. The biological information provided by wearable devices can be used to study happiness from more dimensions; 2. Evaluation indicators will be more inclined to be combined with task construction, which will be more diversified, rationalized and more accurate than existing psychological indicators to describe the degree of happiness, and provide indicators for other tasks related to happiness.

5 Evaluation and data resources

In the research of computational personality, with the popularity of social media and the increase of data collection tools, there are increasing resources available. The evaluation and the data resources for personality prediction, depression detection, suicide detection and happiness evaluation are presented as follow.

5.1 Related evaluation and data resources for personality prediction

5.1.1 (MBTI) Myers-Briggs Personality Type Dataset

There is an open source MBTI indicator-based competition [4] in Kaggle, whose goal is to predict a person's MBTI personality type. The data given is from the Personality Cafe website forum of the user's social media posts. It contains more than 8600 person's information. The first column is the MBTI indicator (4 letters), and the second column is the user's most recent post 50 text messages. The format is shown in Table 9 (only the first 5 data are listed):

Table 9 Myers-Briggs Personality Type Dataset

	Personality	Post
0	INFJ	'http://www.youtube.com/watch?v=qsXHcwe3krw|||...
1	ENTP	'I'm finding the lack of me in these posts ver...
2	INTP	'Good one _____ https://www.youtube.com/wat...
3	INTJ	'Dear INTP, I enjoyed our conversation the o...
4	ENTJ	'You're fired.||| That's another silly misconce...

Participants need to be evaluated into four personality dimensions(each dimension has two types) based on MBTI: extroverted and introverted (E/I), intuition and sensing (N/S), thinking and feeling (T/F), judging and perceiving (J/P). The four dimensions are combined together to describe a personality of a user.

5.1.2 MyPersonality and Twisty

MyPersonality[5] is a Facebook application that allows its users to participate in psychological research by filling out a personality questionnaire. It also provides them with feedback about the score. It was created by David Stillwell in 2007. In 2018, due to the heavy workload of maintaining the dataset, reviewing projects, responding to inquiries and other issues, and complying with various regulations (involving user privacy issues), data is not open to the public any more.

Ben et al. (2016) built a corpus based on Twitter using the MBTI scale, which is available for six languages, as shown in Table 10. It is marked with gender, personality and other tags, a total of 18,168 data.[6]

Table 10 Statistics data for each language (Avg: average number of tweets sent by users)

	# Authors	# Tweets	Avg.	% in-lang
(Avg: average number of tweets sent by users)				
German	411	952,549	2,318	74.9
Italian	490	932,785	1,904	70.6
Dutch	1,000	2,083,484	2,083	74.0
French	1,405	2,786,589	1,983	71.6
Portuguese	4,090	8,833,132	2,160	71.9
Spanish	10,772	18,547,622	1,722	72.8

4 https://www.kaggle.com/c/edsa-mbti/overview.

5 https://sites.google.com/michalkosinski.com/mypersonality.

6 https://www.uantwerpen.be/en/research-groups/clips/research/datasets/.

Table 11 Distribution of depression detection dataset under Twitter

Dataset	Depression	Non-depression	Depression-oriented
User	1402	>3 Billions	36993
Tweets	292564	>100 Billions	35076677

5.2 Related evaluation and data resources for depression detection

5.2.1 Depression detection dataset based on Twitter

Unlike acquiring datasets through time-consuming and laborious interviews and questionnaires, Shen et al. (2017) was inspired by Malmasi et al. (2016), collected data from Twitter. And the corresponding rules determine whether the post from Twitter is indicative of depression. As shown in Table 11, it consists of a depression dataset, a non-depression dataset and a depression-oriented dataset. The depression dataset consists of related depressive articles (292,564 tweets) collected within one month and related 1,402 depressive users; The non-depression dataset is composed of standard active users on Twitter and the tweets they send; because the sample from the depression dataset is too sparse, the difficulty for selecting depression data have increased. And if there were expressions such as "frustration", these would be distributed into the dataset of depression-oriented. The detail of the distribution of the dataset is shown in Table 11.

Nowadays, these datasets can be used for depression detection tasks, and the accuracy rate on this dataset can reach about 85%.

5.2.2 RSDD dataset for depression detection based on Reddit

Yates et al. (2017) collected the self-diagnosis depression detection (RSDD) dataset[7] from the Reddit dataset.[8] The dataset is collected from users who self-diagnosed as depression, while ignoring those users who did not self-diagnose depression. The collected diagnostic reports were labeled by three non-professionals to judge whether the users are suffering from depression. This aimed to keep the mislabeling of data such as "If I am depressed" away from dataset. In addition, due to the user's anonymity, it cannot be confirmed Whether these self-diagnosed reports of depression are real.

This dataset can be used for depression detection, and the accuracy on this dataset can reach 59%.

5.3 Related evaluation and data resources for suicide detection

The shared task for the 2019 workshop on Computational Linguistics and Clinical Psychology Symposium (CLPsych'19) "Predicting the Degree of Suicide Risk in Reddit Posts" [9] is about detecting suicide in social media. It is based on the information published by social media to assess the suicide risk of users. The data comes from Reddit. The three subtasks are all classification tasks with four types, labeled as None, Low, Medium and High suicide risk. The data is divided into two parts, one is from posts in the Reddit suicide section, and the other is from posts in the non-suicide section. The first task only uses posts in the suicide section with a small amount of data; the second task uses posts in the suicide and non-suicide sections; the third task uses only the user's daily posts (non-suicide sections). The specific data and results are in Table 12.

5.4 Relevant evaluation and resources of happiness assessment

The evaluation task comes from "AffCon2019: The 2nd AAAI Workshop on Affective Content Analysis[10]", which is a workshop in AAAI 2019. The dataset used in the evaluation task embraces 10506 data manually marked from the dataset HappyDB and 72324 unmarked data[11]. Each piece of data includes the text of a happy moment described by the users, as well as the time and related tags. The organizer manually marked two new tags: agent and social. The agent describes whether this happy moment is under user's control, and the social describes whether this moment involves anyone other than the user.

The evaluation consists of two tasks: The title of task one is "What are the elements of happiness". It is a semi-supervised learning task based on labeled and unlabeled training data. In the evaluation results, the top three teams are selected and uploaded the model and classification results of each team. The results are shown in Table 13.

The topic of task two is "How can we shape happiness". It is an unsupervised task that proposes new features and insights (unlimited topics) for happy moments in the test set, such as emotions, participants, and content. Task 2 is an open task. There is no unified evaluation standard. The participating teams have conducted further analysis and visualization on the basis of task 1, such as exploring the dependence between the agent and social tags.

7 http://ir.cs.georgetown.edu/resources/rsdd.html.

8 https://files.pushshift.io/reddit/.

9 https://clpsych.org/shared-task-2019-2/.

10 https://sites.google.com/view/affcon2019/cl-aff-shared-task.

11 https://doi.org/10.7910/DVN/JZAS66.

Table 12 Introduction of Suicide Evaluation Data

Task	Source	User scale	Post size	Highest score(macro-f1)
1	Reddit (Suicide section)	1242	1105	0.481
2	Reddit (Suicide and non-suicide sections)	1242	66625	0.457
3	Reddit (Non-suicide section)	1242	70327	0.268

Table 13 Task 1 results

System	Agent		Social	
	Accuracy	F1	Accuracy	F1
ELMo + LSTM + Attention Rajendran et al. (2019)	0.85	0.90	0.92	0.93
Word Pair Convolutional Model (WoPCoM) Saxon et al. (2019)	0.85	0.89	0.91	0.92
inductive transfer learning technique (ITL) Syed et al. (2019)	0.84	0.89	0.92	0.93

6 Interpretablity and Ethical issues

As the research of computational personality begins to use the text and other information of social media, machine learning and deep learning algorithms are more and more widely used. However, compared with the psychological scale, although the applicability of deep learning algorithm is better, it also has interpretability and ethical issues. Firstly, as a black-box algorithm, deep learning algorithm has no good psychological theoretical support and is difficult to explain its results. This hinders the development of relevant algorithms and makes it difficult for models to deeply understand what is personality. Then, it also makes the results of the models more uncontrollable, which makes it difficult for us to avoid some potential risks (such as whether the model is biased against different groups). Secondly, due to the deep learning algorithm need bunch of data, it has been facing the problem of data privacy (Holzinger et al. 2018). How to ensure the protection of privacy for user in the acquisition and use of the data are very controversial issues

Some works have given their answers to these problems. Muller et al. (2021) proposed the ten commands as a reference. However, the model interpretability and ethical issues in the research of computational personality are still urgent to be solved, which is a field worthy of research.

7 Conclusion

In this paper, we review recent advances and available data resources in computational personality in four aspects: personality prediction, depression detection, suicide detection, and happiness assessment. We have also conducted an in-depth discussion on related fields. In particular, we construct the overall research framework of computational personality, which provides an overview of the research from resource to application. Furthermore, we discuss the ethic issue and interpretability of the algorithms. We hope that our work will encourage further interdisciplinary research on computational personality and facilitate progress in this area.

Funding This work was supported by the Natural Science Foundation of China (61702080, 61632011, 62076046, 62006130, 61976036), and Major science and technology projects of Yunnan Province (202002ab080001-1).

Availability of data and material The authors make sure that all data and materials support their published claims and comply with field standards.

Code Availability The authors make sure that custom code support their published claims and comply with field standards.

Declarations

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

References

Adelstein JS, Shehzad Z, Mennes M, DeYoung CG, Zuo X-N, Kelly C, Margulies DS, Bloomfield A, Gray JR, Castellanos XF, Milham MP (2011) Personality is reflected in the brain's intrinsic functional architecture. PLoS ONE 6(11):1–12

Amini Payam, Ahmadinia Hasan, Poorolajal Jalal, Amiri Mohammad Moqaddasi (2016) Evaluating the high risk groups for suicide: A comparison of logistic regression, support vector machine, decision tree and artificial neural network. Iranian J Public health 45(9):1179

An M, Wang J, Li S, et al (2020) Multimodal topic-enriched auxiliary learning for depression detection[C]//proceedings of the 28th international conference on computational linguistics. 1078-1089

Asai, A., Evensen, S., Golshan, B., Halevy, A., Li, V., Lopatenko, A., Stepanov, D., Suhara, Y., Tan, W.C., Xu, Y (2018) Happydb: a corpus of 100,000 crowd sourced happy moments. In: Proceedings of LREC 2018. European language resources association (ELRA), Miyazaki, Japan (May 2018)

Bagge Courtney, Osman Augustine (1998) The suicide probability scale: norms and factor structure. Psychol Rep 83(2):637–638

Bagroy S, Kumaraguru P, and Choudhury MD (2017) A social media based index of mental happiness in college Campuses. In Proceedings of the 2017 CHI conference on human factors in computing systems (CHI'17). Association for computing machinery, New York, NY, USA, 1634-1646. https://doi.org/10.1145/3025453.3025909

Beck AT, Ward CH, Mendelson M et al (1961) An inventory for measuring depression[J]. Archives Gen Psychiatry 4(6):561–571

Bhat HS and Goldman-Mellor SJ (2017) Predicting adolescent suicide attempts with neural networks. arXiv preprint arXiv:1711.10057

Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation[J]. J Mach Learn Res 3:993–1022

Bradley M M, Lang P J (1999) Affective norms for English words (ANEW): instruction manual and affective ratings[R]. Technical report C-1, the center for research in psychophysiology, University of Florida

Burnap Pete, Colombo Gualtiero, Amery Rosie, Hodorog Andrei, Scourfield Jonathan (2017) Multi-class machine classification of suicide-related communication on twitter. Online Soc Netw Media 2:32–44

Cao L, Zhang H, Feng L, Wei Z, Wang X, Li N, He X (2019) Latent suicide risk detection on microblog via suicide-oriented word embeddings and layered attention. EMNLP/IJCNLP (1) : 1718-1728

Capraro R, Capraro M (2002) Myers-Briggs type indicator score reliability across studies: a meta-analytic reliability generalization study. Edu Psychol Meas 62(4):590–602

Celli F, Bruni E, Lepri B (2014) Automatic personality and interaction style recognition from facebook profile pictures. MM '14: Proceedings of the 22nd ACM international conference on multimedia november 2014 Pages 1101-1104

Chen X, Sykora MD, Jackson TW et al (2018) What about mood swings: identifying depression on twitter with temporal measures of emotions[C]. Companion proceedings of the the web conference:1653–1660

Chiranjeevi, Rahul V, Elangovan D (2019) "A review on human action recognition and machine learning techniques for suicide detection system." Adv Intell Syst Comput , vol 939

Choudhury MD, Kiciman E, Dredze M, Coppersmith G, Kumar M (2016) Discovering shifts to suicidal ideation from mental health content in social media. In Proceedings of the 2016 CHI conference on human factors in computing systems, pages 2098-2110. ACM

Christian H, Suhartono D, Chowanda A et al (2021) Text based personality prediction from multiple social media data sources using pre-trained language model and model averaging[J]. J Big Data 8(1):1–20

Cloninger CR, Svrakic DM, Przybeck TR (2006) Can personality assessment predict future depression? A twelve-month follow-up of 631 subjects[J]. J Affect Disord 92(1):35–44

Coppersmith, G et al (2016) Exploratory analysis of social media prior to a suicide attempt. In: Proceedings of the third workshop on computational linguistics and clinical psychology

De Choudhury M, Gamon M, Counts S, et al (2013) Predicting depression via social media[C]. Seventh international AAAI conference on weblogs and social media

Diener E (1984) Subjective happiness. Psychology Bulletin. 95 (2)

DieNer E, Suh EM, Lucas RE, Smith HL (1999) Subjective happiness: three decades of progress. Psychol Bull 125:276–302

Dodds PS, Harris KD, Kloumann IM, Bliss CA, Danforth CM (2011) Temporal patterns of happiness and information in a global social network: hedonometrics and twitter. PLoS ONE 6(12):e26752

Du J, Zhang Y, Luo J, Jia Y, Wei Q, Tao C, Xu H (2018) Extracting psychiatric stressors for suicide from social media using deep learning. BMC Med Inf Decis Making 18(2):43

Funder DC (1995) On the accuracy of personality judgment:a realistic approach[J]. Psychol Rev 102(4):652

Funder DC (2001) Personality[J]. Annual Rev Psychol 52(1):197–221

Gaur Manas, Alambo Amanuel, Sain Joy Prakash, Kursuncu Ugur, Thirunarayan Krishnaprasad, Kavuluru Ramakanth, Sheth Amit, Welton Randy, Pathak Jyotishman (2019) Knowledge-aware Assessment of Severity of Suicide Risk for Early Intervention. The world wide web conference (WWW'19). Association for Computing Machinery, New York, NY, USA, pp 514–525

Gjurković M, Šnajder J. Reddit: a gold mine for personality prediction. Proceedings of the second workshop on computational modeling of people's opinions, personality, and emotions in social media, pages 87-97

Golbeck J, Robles C, Turner K (2011) Predicting personality with social media[C]. CHI'11,

Gui T, Zhu L, Zhang Q et al (2019) Cooperative multimodal approach to depression detection in twitter[C]. Proceedings of the AAAI conference on artificial intelligence. 33:110–117

Gupta G K, Sharma D K (2021) Depression detection on social media with the aid of machine learning platform: a comprehensive survey[C]//2021 8th international conference on computing for sustainable global development (INDIACom). IEEE,658-662

Halawa MS, Shehab ME, Hamed EMR (2015) Predicting student personality based on a datadriven model from student behavior on LMS and social networks. IEEE

Hao B., Li L., Gao R., Li A., Zhu T (2014) Sensing subjective happiness from social media. In: Ślezak D., Schaefer G., Vuong S.T., Kim YS. (eds) Active media technology. AMT 2014. Lecture Notes in Computer Science, vol 8610. Springer, Cham

Harris Keith M, Syu Jia-Jia, Lello Owen D, EileenChew YL, Willcox Christopher H, Ho Roger HM (2015) The abcs of suicide risk assessment: applying a tripartite approach to individual evaluations. PLoS One 10(6):e0127442

Hiraga M (2017) Predicting depression for japanese blog text[C]. Proceedings of ACL 2017, Student Research Workshop. 107-113

Holzinger, Andreas, Kieseberg P, Weippl ER, Tjoa AM (2018) Current advances, trends and challenges of machine learning and knowledge extraction: from machine learning to explainable AI. CD-MAKE

Ilmini W.M.K.S , Fernando T.G.I (2018) Computational personality traits assessment: A review[C]// 2017 IEEE International conference on industrial and information systems (ICIIS). IEEE,

Jashinsky, Jared Michael, Scott H. Burton, Carl Lee Hanson, Joshua H. West, Christophe G. Giraud-Carrier, Michael D Barnes ,Trenton Argyle (2014) Tracking suicide risk factors through twitter in the US. Crisis 35 1 : 51-9

Jones Natalie J, Bennell Craig (2007) The development and validation of statistical prediction rules for discriminating between genuine and simulated suicide notes. Arch Suicide Res 11(2):219–233

Kalchbrenner N, Grefenstette E, Blunsom P. A (2014) convolutional neural network for modelling sentences[J]. arXiv preprint arXiv:1404.2188,

Kampman O, Barezi Elham J, Bertero D, Fung P (2018) Investigating audio, video, and text fusion methods for end-to-end automatic personality prediction. ACL 2:606–611

Kang K, Yoon C, Kim E Y (2016) Identifying depressive users in twitter using multimodal analysis[C]. 2016 International conference on big data and smart computing (BigComp). IEEE, : 231-238

King-wa Fu, Liu Ka Y, Yip Paul SF (2007) Predictive validity of the chinese version of the adult suicidal ideation questionnaire: Psychometric properties and its short version. Psychol Assess 19(4):422

Kosinski Michal, Stillwell David, Graepel Thore (2013) Private traits and attributes are predictable from digital records of human behavior. Proc the Nat Acad Sci 110(15):5802–5805
Lei Liu, Wujun Sun, Yuan Jiang, Ping Fang (2019) Evaluation and prospect of happiness measurement index system[J]. China Spec Ed 02:66–73
Li, Jun, Yan Z, Lin Z, Liu X, Leong HV, Yu NX, Li Q (2021) Suicide ideation detection on social media during COVID-19 via adversarial and multi-task Learning. APWeb/WAIM
Lipeng Xu, Wenai Song (2019) Suicide idea detection based on Chinese microblog language features [J]. J North Univ China (Nat Sci Edition) 40(04):350–357
Liu P, Cui Z, Zhou W, Zhang Y (2019) Research on the character prediction of Weibo users based on behavior information. J Beijing Univ Inf Sci Technol, Vol 34(3)
Lynn V, Balasubramanian N, Schwartz H A (2020) Hierarchical modeling for user personality prediction: the role of message-level attention[C]//Proceedings of the 58th annual meeting of the association for computational linguistics. 5306-5316
Malmasi S, Zampieri M, Dras M (2016) Predicting post severity in mental health forums[C]. Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology. : 133-137
Moreno MA, Jelenchick LA, Egan KG et al (2011) Feeling bad on facebook: depression disclosures by college students on a social networking site[J]. Depress Anxiety 28(6):447–455
Muller H, Mayrhofer M, Van Veen E, Holzinger A (2021) The ten commandments of ethical medical AI. Computer 54(07):119–123
Oxman T E, Rosenberg S D, Tucker G J (1982) The language of paranoia[J]. Am J Psychiatry
Park M, Cha C, Cha M (2012) Depressive moods of users portrayed in twitter[C]. Proceedings of the ACM SIGKDD workshop on healthcare informatics (HI-KDD). 2012, : 1-8
Park M, McDonald D W, Cha M (2013) Perception differences between the depressed and non-depressed users in twitter[C]. Seventh international AAAI conference on weblogs and social media
Pedregosa F, Varoquaux G, Gramfort A et al (2011) Scikit-learn: machine learning in Python[J]. J Mach Learn Res 12:2825–2830
Pennebaker JW, King LA (1999) Linguistic styles: language use as an individual difference[J]. J Pers Soc Psychol 77(6):1296
Pestian J, Nasrallah H, Matykiewicz P, Bennett A, Leenaars A (2010) Suicide note classification using natural language processing: a content analysis. Biomedical Inform Insights, 3:BII-S4706
Pianesi F (2013) Searching for personality [social sciences] [J]. IEEE Signal Process Mag 30(1):146–158
Qi Jiayin, Xiangling Fu, Zhu Ge (2015) China subjective happiness measurement based on Chinese grassroots blog text sentiment analysis. Inf & Manag 52:859–869
Rajendran A, Zhang C, Abdul-Mageed M (2019) Happy together: Learning and understanding appraisal from natural language. In: Proceedings of the 2nd workshop on affective content analysis @ AAAI (AffCon2019). Honolulu, Hawaii (January 2019)
Resnik P, Garron A, Resnik R (2013) Using topic modeling to improve prediction of neuroticism and depression in college students[C]. Proceedings of the 2013 conference on empirical methods in natural language processing. : 1348-1353
Ringeval F, Schuller B, Valstar M, et al (2017) Avec 2017: Real-life depression, and affect recognition workshop and challenge[C]. Proceedings of the 7th annual workshop on audio/visual emotion challenge. : 3-9
Rohan Mishra, Pradyumna Prakhar Sinha, Ramit Sawhney, Debanjan Mahata, Puneet Mathur, Rajiv Ratn Shah (2019) SNAP-BATNET: cascading author profiling and social network graphs for suicide ideation detection on social media. NAACL-HLT (Student research workshop) : 147-156
Rude S, Gortner EM, Pennebaker J (2004) Language use of depressed and depression-vulnerable college students[J]. Cognit & Emot 18(8):1121–1133
Samareh A, Jin Y, Wang Z, et al (2018) Predicting depression severity by multi-modal feature engineering and fusion[C]. Thirty-second AAAI conference on artificial intelligence
Sawhney, Ramit, Harshit Joshi, Rajiv Ratn Shah, Lucie Flek (2021) Suicide ideation detection via social and temporal user representations using hyperbolic learning. NAACL
Sawhney, Ramit, Joshi H (2021) PHASE: learning emotional phase-aware representations for suicide ideation detection on social media. EACL
Sawhney, Ramit, Joshi H, Gandhi S and Shah RR (2020) A time-aware transformer based model for suicide ideation detection on social media. EMNLP
Saxon M, Bhandari S, Ruskin L, Honda G (2019) Word pair convolutional model for happy moment classification. In: Proceedings of the 2nd workshop on affective content analysis @ AAAI (AffCon2019). Honolulu, Hawaii (January 2019)
Shen G, Jia J, Nie L, et al (2017) Depression detection via harvesting social media: a multimodal dictionary learning solution[C]. IJCAI.: 3838-3844
Shen T, Jia J, Shen G, et al (2018) Cross-domain depression detection via harvesting social media[C]. International joint conferences on artificial intelligence,
Shlomo A, Sushant D, Moshe K, James W. Pennebaker (2005) Lexical predictors of personality type. In Proceedings of the 2005 joint annual meeting of the interface and the classification society of North America
Singh, Kuldeep & Shakya, Harish & Biswas, Bhaskar (2017) Happiness index in social network. https://doi.org/10.1145/3025453.3025909.
Skowron M, Tkalcic M, Ferwerda B, Schedl M (2016) Fusing social media cues: personality prediction from twitter and instagram. WWW (Companion Volume) : 107-108
Sueki Hajime (2015) The association of suicide-related Twitter use with suicidal behaviour: a cross-sectional study of young internet users in Japan. J Affect Disord 170(2015):155–160
Süleyman Aslan, Uğur Güdükbay. Multimodal Video-based Apparent Personality Recognition Using Long Short-Term Memory and Convolutional Neural Networks. Computer Vision and Pattern Recognition (cs.CV)
Syed B, Indurthi V, Shah K, Gupta M, Varma V (2019) Ingredients for happiness: modeling constructs via semi-supervised content driven inductive transfer learning. In: proceedings of the 2nd workshop on affective content analysis @ AAAI (AffCon2019). Honolulu, Hawaii (January 2019)
Thelwall M, Buckley K, Paltoglou G et al (2010) Sentiment strength detection in short informal text[J]. J Am Soc Inf Sci Technol 61(12):2544–2558
Tim MH Li, Ben CM Ng, Michael Chau, Paul WCWong, Paul SF Yip (2013) Collective intelligence for suicide surveillance in web forums. In Pacific asia workshop on intelligence and security informatics, pages 29-37. Springer
Verhoeven B, Daelemans W, Plank B (2016) TWISTY: a Multilingual twitter stylometry corpus for gender and personality profiling
Walsh Colin G, Ribeiro Jessica D, Franklin Joseph C (2017) Predicting risk of suicide attempts over time through machine learning. Clin Psychol Sci 5(3):457–469
Wang T, Brede M, Ianni A, et al (2017) Detecting and characterizing eating-disorder communities on social media[C]. Proceedings of the Tenth ACM international conference on web search and data mining. : 91-100
Wilson T, Hoffmann P, Somasundaran S, et al (2005) OpinionFinder: a system for subjectivity analysis[C]. Proceedings of HLT/EMNLP 2005 interactive demonstrations.: 34-35

Xiaolei Huang, Xin Li, Tianli Liu, David Chiu, Tingshao Zhu, Lei Zhang (2015) Topic model for identifying suicidal ideation in chinese microblog. Proceedings of the 29th pacific asia conference on language, information and computation, pages 553-562

Xingyu Xu (2017) A summary of subjective happiness [J]. Mod Econ Inf 20:363–364

Xue C, Qi H, Yuxuan L, Shuya Z, Ge Z (2019) Research on user model based on consumer character ontology. 34, (3)

Yamada K, Sasano R, Takeda K (2019) Incorporating textual information on user behavior for personality prediction. ACL (2) : 177-182

Yan Li, Jun Zhao (2004) Overview of research on happiness [J]. J Shenyang Norm Univ (Soc Sci Ed) 02:22–26

Yates A, Cohan A, Goharian N (2017) Depression and self-harm risk assessment in online forums[J]. arXiv preprint arXiv:1709.01848,

You S, DesArmo J, Joo S (2013) Measuring happiness of US cities by mining user-generated text in Flickr.com: a pilot analysis. In Proceedings of the 76th ASIS&T Annual meeting: beyond the Cloud: rethinking information boundaries (ASIST'13). Am Soc Inf Sci, USA, Article 167, 1-4

Zafar A, Chitnis S (2020) Survey of depression detection using social networking sites via data mining[C]//2020 10th International conference on cloud computing, data science & engineering (confluence). IEEE, 88-93

Zheng H, Zuo W (2016) Multi-label social network user personality prediction based on information gain and semantic features. J Jilin Univ. Vol 54 3

Zogan H, Razzak I, Jameel S, et al (2021) DepressionNet: learning multi-modalities with user post summarization for depression detection on social media[C]//proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval. 133-142