IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 4, 2018, accepted July 15, 2018, date of publication July 31, 2018, date of current version September 21, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2859440

A Hybrid User Experience Evaluation Method for Mobile Games

QI YU[1], XIAOPING CHE[1], (Member, IEEE), SIQI MA[1], SHIRUI PAN[2], YUXIANG YANG[1], WEIWEI XING[1], AND XIMENG WANG[2,3]

[1]School of Software Engineering, Beijing Jiaotong University, Beijing 100044, China
[2]Centre for Artificial Intelligence, Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, NSW 2007, Australia
[3]School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China

Corresponding author: Xiaoping Che (xpche@bjtu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61502028 and in part by the Fundamental Research Funds for the Central Universities under Grant 2018JBM073.

ABSTRACT With the development of the mobile phone industry, mobile applications market becomes thriving. However, the immature information technology and unfriendly interface bring negative user experience (UX) to the mobile users and thus affect the service life of mobile applications, especially for the most concerned entertaining applications, mobile games. As a result, the evaluating UX and finding crucial factors of UX become a challenge. Over the last decades, numerous researches have tried to deal with this issue, but none of them has clearly identified the relations among positive-negative UX, sufficient human characteristics and specific events of applications. This paper proposes a subjective-objective evaluation method. Subjective UX of mobile games is sufficiently obtained and objective UX is verified through the electrocardiogram signals and heart rate variability. In order to reveal distinct relations among UX factors, sufficient user characteristics and categories of game events, an improved RIPPER algorithm is proposed by this paper to obtain the relations. Experiments are performed with 300 testers who played mobile parkour games for at least five minutes. The accuracy and efficiency of the proposed method have been verified through real experiments and objective measures. In addition, this paper provides an effective sampling method and a data analysis algorithm to obtain crucial UX factors for mobile applications.

INDEX TERMS User experience, data analysis, mobile games, ripper, HRV.

I. INTRODUCTION

Digital mobile technology is growing day by day, and the revolution of mobile phones brings more promising prospects. Large number of mobile phone applications have been developed successfully. Among all the applications, entertainment activities are more attractive to mobile users, particularly mobile games. The number of mobile phone users grows to 1.86 billion by the end of 2015, and they averagely take 20% of their leisure times on playing mobile games [1]. As a result, with such a huge market, the average cost of each user for mobile games rise to 40 dollars per device by the end of 2016 [2].

Among all the software features, User Experience (UX) is the most important factor to reflect the sense of mobile users when using mobile applications. Thus it would significantly affect the application's life cycle, especially for mobile games [3], [4]. Therefore, keeping a high level of UX is crucial for extending the service life of mobile applications.

Although the evaluation of UX has been applied for mobile applications for years, there are still many challenges to accurately evaluate and effectively improve UX. Hardware difference, interface appearance, friendliness of application and different event category would also impact UX [5]. What is more, poor configuration of the above factors would give negative experience to users. As a result, it is crucially important to recognize the factors which would actually affect UX.

In recent years, numbers of researches have explored this issue. Existing methods of assessing UX include subjective methods, objective methods, and subjective-objective methods. Subjective methods are mainly based on user self-evaluation, such as verbal and questionnaire assessment. They can evaluate the direct feeling of different aspects of products and use learning technique to rank aspects [6], [7]. Objective methods mainly rely on user's physiological signal, such as brain waves, heartbeats, skin conductivity, respiration, etc [8]. Since physiological signal is continuous but subjective experience is partial, objective methods have higher temporal resolution compared with subjective methods [9]. However, subjective evaluation or objective evaluation is still one-sided. In recent years, a large number of studies focus on

2169-3536 © 2018 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

the combination of subjective and objective methods. In the combined methods, questionnaire can effectively evaluate user's subjective experience on different aspects of products [10], and physiological signal can effectively measure user's exciting point [12], thus the combined methods provide more comprehensive and convincing UX evaluation. Different factors that affect UX are chosen and tested from the three kinds of methods mentioned above.

However, current studies of mobile game UX still lack sufficient factors and none of the studies has explored the relations among user characteristics, categories of game events and positive-negative UX [3]. The purpose of this paper is to precisely evaluate UX by combining subjective method with objective method, in order to obtain sufficient specific factors that influence UX. Besides, an improved RIPPER algorithm proposed by this paper is used to find out the consequences among user characteristics, specific game events and UX. These consequences are represented in the form of rule sets. Electrocardiogram (ECG) signals and heart rate variability (HRV) of testers are measured as objective factors to verify the consequences. The rule sets can precisely locate the user characteristics and type of events that have positive or negative influence on UX. This conclusion can also provide a reasonable suggestion for improving the UX quality of games by resetting corresponding events according to the targeting users, and consequently increases the diffusion and service life of games.

In this paper, a popular mobile game 'Temple Run' is used as the testing game and the experiment is performed on mobile phones with mainstream hardware. User's subjective experience and characteristics are obtained through designed questionnaires. The game events are redefined by reviewing existing studies and recorded by camera. The user's ECG data is collected through professional instrument during the playing process. All data is pre-processed to obtain digitized value, and then values are analyzed by the improved RIPPER algorithm to obtain the rule sets that relate to UX.

The paper's primary contributions are:

- A subjective method that includes UX assessment and sufficient user characteristics related to UX is designed.
- An improved RIPPER algorithm is proposed to deal with missing samples with specific attribute values.
- The relations among user characteristics, categories of game events and positive-negative UX are obtained by the improved RIPPER algorithm, and objectively verified by HRV data.

The rest of the paper is organized as follows: Section II presents a brief review of assessing UX methods that include subjective method, objective method and subjective-objective method. Section III presents the definition of user characteristics in the questionnaire and game events. Section IV describes the data pre-processing method and the improved RIPPER algorithm in detail. Section V reports the experimental results. Finally, the work of this paper is concluded and future directions are provided in Section VI.

II. RELATED WORKS

As mentioned in Section I, UX is the crucial factor for extending service life of games. However, a precise evaluation of UX is a challenging task. In recent years, lots of researches have paid their attentions to the subjective methods, objective methods or subjective-objective methods.

A. SUBJECTIVE METHODS

The subjective method can evaluate the direct UX of various aspects of the product. Therefore, a great deal of researches have used questionnaire to evaluate user's subjective experience.

Arttu Perttula conducts two serious game tests on 102 high school students [14]. Subjective UX is obtained by flow questionnaire so as to find out the factors that affect players and improve the game. The results show that the flow experience can be used to evaluate the overall quality of the game and provide a structured way to evaluate the quality of the game. However, the flow experience does not provide detail information about the weaknesses of the game, thus a complementary approach to find the cause is needed. Martin Schrepp designs a questionnaire that includes several typical questions related to the subjective UX [15]. Katy Tcha-Tokey analyzes the existing questionnaires that relate to education and entertainment in the field of Virtual Reality (VR), and then creates a universal template of questionnaire for most areas of VR [16]. Martin Schrepp creates a UX questionnaire in Portuguese and a children's special version to adapt to the language comprehension of the intended target group. The experiment results show that the method can effectively evaluate vary levels of UX [10]. The methods mentioned above all design questionnaire in related fields by reviewing existing literature. However, the relation among UX and user's own characteristics and detailed product factors are still missing. The researches are all based on statistics and the results have not been verified.

A large number of researches analyze factors related to subjective UX through machine learning methods. In [19], Bernhaupt R finds out the characteristics that relate to UX according to the existing relevant literature and develops a series of attributes about the interactive television to evaluate UX. The maximum likelihood factor analysis method is used to analyze the dimensions of the questionnaire, that is, beauty, emotion, stimulation and identification are used to evaluate the UX of the interactive television (iTV) systems. Tcha-Tokey et al. [20] survey 253 enterprise resources planning (ERP) users by questionnaire and analysis data of the questionnaires by hierarchical multiple regression. The results show that user adaptability has a strong positive impact on ERP users and appropriate strategies should be formulated to effectively manage adaptability of ERP user. Reference [2] studies the relations among user characteristics, game features and intention to download mobile games. User characteristics includes gender and positive-negative emotions. Game features includes application price, game content quality and game time. Data from 535 mobile users are analyzed by fuzzy-set qualitative analysis (fsQCA). Ten intentions that explain highly downloads are obtained. However, the above methods only find part of human characteristics and product characteristics related to UX. The detail features of product and other human characteristics are not considered.

As mentioned above, user's subjective UX obtained by questionnaire is not verified. Thus, inaccurate UX will affects the exploration of relevant factors.

B. OBJECTIVE METHODS

Physiological changes plays an important role in emotional experience. Many researches use physiological signals to objectively evaluate UX. Healey JA measures ECG, electromyography (EMG), skin conductance and respiration of 24 drivers. The test phases include rest time, driving on highway and driving in the city, which represent different levels of stress. The results show that skin conductivity and ECG positively relate to driver stress levels [8]. Seokbin Kang develops software that integrates real-time physiological sensing, whole body interaction, and visually responsive large screens. The software changes the responsive large screen through the user's physiological response so as to facilitate interaction. This experiment tests twenty teachers and the results show that close coupling among physical interaction, perception and visualization helps to promote interaction, support and promote of social activities, and increases the experience of interests [21]. [22] studies the physiological characteristics of bus drivers in the parking process. Physiological data such as BSA and ECG of the driver are compared under real vehicle test with natural condition. The results show that the driver in the bus travel more sensitive than the natural state because of the external environment (platform settings, traffic flow, etc.). The above methods all use statistical methods to analyze the relation between physiological characteristics and UX. However, physiological characteristics are different with different user, thus user characteristics should also be considered.

Also, a large number of researches have analyzed factors related to objective UX through machine learning methods. In [23], Blood Volume Pressure (BVP), GSR and EMG are collected by three sensors and analyzed by decision tree classifier. Experimental results show that the accuracy of the method is 91%. For improving the performance of classification, [24] and [25] establish an automatic recognition system based on supervised machine learning. References [26] and [27] propose classification systems based on graph feature selection. Physiological data (include breath, ECG and skin conductance) from 24 participants are collected. The results show that the accuracy of the model is 70%. The purpose of [28] is to find the relations between changes of multi-modal body signal and discrete emotional states. Perez-Rosero MS establishes an individual-based inference model and analyzes EMG, BVP and Galvanic Skin Response(GSR) in eight emotional states by this model. The results show that the accuracy of emotion recognition by multi-signal model is 88.1%. Some other learning based and graph based works also contribute to this domain [11], [13], [17], [18]. Nevertheless, the existing objective methods only evaluate UX by physiological changes and do not consider the relation between user characteristics and UX.

Objective UX are obtained by analysing physiological signal on different states. However, the states defined by person are not precise and the physiological responses of person with different characteristics in the same state are completely different.

C. SUBJECTIVE-OBJECTIVE METHODS

The quality of user experience can be evaluated more accurately by combining physiological signals with questionnaires [29]. The questionnaire can effectively evaluate user's subjective experience on different scales of products, and physiological signals can reflect user's mood objectively.

A lot of researches combine objective with subjective methods. Liapis and Xenos [30] implements a software that allows human-computer interaction. He evaluates user's emotional experience using heart rate, galvanic skin response, and blood volume pressure when interacting. Then the objective results compared with the subjective UX filled in the questionnaire by testers. In [31], the testers watch the pictures and verbally answer like or dislike. At the same time, brainwaves that objectively reflect the testers' emotions are recorded through the eMotive earphones when they watching a specific picture. Then objective emotions contrast the real emotions expressed by the pictures. In [29], testers play computer game with four different difficult stages: beginning, easy, medium, hard. At the same time, heartbeat is measured during playing. After each stage, testers score for 4 experience that includes boring, challenging, frustration and fun on the questionnaire. The range of score is 1 to 5, which stand for different scales of 4 experience. The results show that players who played the game for the first time have higher heartbeat in beginning than medium and hard stages, while experienced players have faster heartbeat in hard stage than in beginning and easy stages. In order to investigate audiences' experience when they participating in the movie, audiences are asked to participate in two versions of the 3D virtual movie. One is that audience plays a role alone, the other is that audience plays with friends. After participating, audiences are asked to watch original movie and movie that they participated. Meanwhile, their skins are measured. The results show that audiences have a subjective sense of participation and emotional reaction, and the GSR of the participating audiences increased in stage [32]. The above methods all measure the physiological responses of different UX and compare them with subjective experiences. However, the above methods do not consider user characteristics and detailed event factors related to user experience.

Some researches study the relation between user experience and product events. In [12], physiological responses, task performance, and self-report data of participants are collected and analyzed. The study shows that physical indicators varied with task performance, and participants have greater changes of GSR in failure tasks. In addition, there is also a relations between GSR and UX self-reported. To investigate the impact of different inputs on the game experience, tester's electroencephalograph (EEG) is measured when testers using a mobile touch controller or a conventional gamepad to play the game. Then each tester's EEG compared with Pragmatic Quality (PQ), Hedonic Quality -Stimulation (HQS), Hedonic Quality - Identity (HQI) and Attractiveness (ATT) on the questionnaire. The study shows that different controllers have different effects to emotions [33]. The above researches consider the impact of failure events and successful events on UX, but do not consider user characteristics such as gender, age, and personality traits. Similarly, the detailed events may affect UX, such as event's category.

Some researches study the relations between factors and UX. [9] interprets ECG data using log analysis and self-assessment. Game behaviors of 55 testers are recorded through the game log when measuring ECG. After playing, game experiences are filled in questionnaire. To obtain the relations between game events and experience, game events are divided into success and fail. To find the factors that affect UX, the questionnaire includes seven dimensions: flow, sensory and imaginative immersion, ability, challenge, positive and negative effects, and stress. Regression analysis is used to analyze the relations among heart rate variability(HRV), the type of game events and number of game events. The results show that the decrease of HRV relate to the higher self-reported score of ''nervous'' during the game. The more actions, the greater deceleration of HRV. Although it considers the effect of fail and success events to UX, but fails to consider the characteristics of user such as gender, age, personality and categories of the game events.

In [6], Arttu Perttula measures the relations between synergistic heartbeat and UX of audience who watch hockey game. In experiment, testers use heart rate band with bluetooth to connect the mobile client and then display the heart rate of each tester on mobile screen. To obtain synergistic heart rate, all mobile clients communicate with the server through wireless and then server displays the collective heart rate of all users at each moment. After game, events are annotated on the heartbeat graph. Then they design a questionnaire that includes interest of heart rate of four events. The user's interests in heart rate of the four events are used as features to evaluate UX and further enhance UX of the public game. The results show that synergistic heart rate has potential value to objective evaluation. However, only four events are considered rather than all events. The user characteristics are also not considered and the relation among UX, user characteristics and game events are not obtained.

In all methods mentioned in Section II, the methods using only subjective or objective methods can not fully assess UX. Existing subjective-objective methods do not consider the relations among UX, sufficient user characteristics and categories of game events. In general, the methods anlyzed the relations between UX and related factors are maximum likelihood factor analysis, regression analysis, fuzzy-set qualitative analysis(fsQCA) and decision tree. This study is a classification problem, regression analysis is not applicable. Maximum likelihood factor analysis reduces dimensionality of multiple factors, which makes the factors not acceptable. FsQCA can obtains the relations between UX and relevant fators, but the values of factors are fuzzy. For example, older person, the older is not clear.

Inspired by these methods, questionnaire and HRV are used to obtain subjective and objective UX. In the work of [9] shows that UX relates to the number of failure and successful events, thus death rates of different events categories are considered as factors that affect the UX. The purpose of this paper is to find out the rule sets that illustrate the relations among positive-negative UX, user characteristics and categories of game events. In this paper, characteristics are expended based on characteristics proposed in [2] to obtain sufficient characteristics and then shown in questionnaire. Game events are defined and classified based on game events proposed in [3] and [9]. Rule learning algorithms and decision tree can both describe relations between UX and relevant factors. Among rule learning algorithm, the RIPPER algorithms is better in noisy data set than decision tree. This study choose the RIPPER algorithm and improved it to adapt to missing samples.

III. DEFINITION

The purpose of this paper is exploring the relations among positive-negative UX, user characteristics and categories of game events. In this section, questionnaire and game events are defined. Sufficient user characteristics and subjective experience are also included in questionnaire.

A. QUESTIONNAIRE

As [2]mentioned, gender and age are factors related to the intention of downloading and playing mobile games. Therefore, they are the primary attributes we considered for user characteristics. For acquiring comprehensive user characteristics profiles, Parkour game experience [6], Physical exercise frequency [29], Character traits [19], Basic heart rate [8], [9], Systolic pressure, Diastolic pressure and Oxygen saturation [15] are all added in the questionnaire.

After determining the attributes, distinct values are used for all attributes. According to the distribution of the game players' ages, the values of age are defined into different sets from 11 to 50 [2]. The values of Parkour game experience are discretized into rich and poor [6], and the values of physical exercise frequency are discretized into frequent and occasional [29]. The values of character traits are discretized into extroverted and introverted [19]. Conventionally, the range of normal heart rate is 60-100, but some person's heartbeat beyond or below this range. Therefore, basic heart rates less than 60 and greater than 100 are discretized into abnormal values [8]. The range of normal systolic pressure is 90-140. Systolic pressures less than 90 and more than 140 are considered as abnormal. Similarly, diastolic pressures and oxygen

TABLE 1. Subjective questionnaire.

Attributes	Value	Value
Age	11-15 16-20 21-25 26-30 31-35 36-40 41-45 46-50	
Gender	male	female
Parkour game experience	rich	poor
Physical exercise frequency	frequent	occasional
Character traits	extroverted	introverted
Basic heartbeat	normal	abnormal
Systolic pressure	normal	abnormal
Diastolic pressure	normal	abnormal
Oxygen saturation	normal	abnormal
First experience of this game	yes	no
Willing to continue playing	yes	no

saturation are also divided into normal set and abnormal set [15]. The range of normal diastolic pressure is 60-90, and normal oxygen saturation is 98%. To evaluate the user's subjective experience of games, user's overall assessment is added in the questionnaire [2]. The questionnaire is shown in Table 1. In the evaluation values, "willing to continue playing" is considered as positive UX, "not willing to continue playing" is considered as negative UX [31].

B. GAME EVENTS

Inspired by that UX relate to death rate of total game events [9], this paper considers the death rates of different events categories. In this section, all game events are described and classified. The game event is defined as a hurdle event that needs to be overcame by user. For example, a player needs to jump ahead to evade the running bear that behind the player to catch player, needs to retreat to avoid the fixed horizontal trees over the ground, needs to jump to avoid the fixed stone or the moving and rotating wheel on the ground. The events are listed below:

(a) The fixed iron chain over the cliff.
(b) The fixed poll on the ground.
(c) The fixed cliff.
(d) The fixed stone on the ground.
(e) The fixed flame over the ground.
(f) The fixed red brick on the ground.
(g) The fixed and rotating wheel on the ground.
(h) The moving and rotating wheel on the ground.
(i) The road with half width of the normal road.
(j) The road with one quarter of the normal road width.
(k) The fixed horizontal trees over the ground.
(l) The fixed flame on the ground.
(m) The fixed and rotating saw over the ground.
(n) The running bear behind to catch the player.

In the events defined above, game events are divided into top and bottom events based on the user's opposite operation like jump or retreat. According to the events are moving or not, they are divided into moving and fixed events. Among the events, (a)-(b) belong to bottom events, (k)-(m) belong to top events. (a)-(g) and (i)-(l) belong to fixed events, (h) and (n) belong to moving events.

TABLE 2. Various categories of game events.

Category of game event	Value	Value
Death rate of fixed game events	high	low
Death rate of moving game events	high	low
Death rate of top game events	high	low
Death rate of bottom game events	high	low

After defining game events, the death rate values of game events in different categories are discretized. Among the above four types of game events, the fixed and the moving events are antitheses, the top and the bottom events are antitheses. Based on the value of two opposing events, the death rates are divided into two categories, high and low. For example, death rate of moving events will be low if death rate of fixed events of the user is high. Similarly, death rate of bottom events will be low if death rate of top events of the user is high. Each user's death rate of fixed game events is counted by $R_{df} = \frac{N_{df}}{N_{tf}}$, where N_{tf} is the total number of fixed events one user experienced, and N_{df} is the total number of dead fixed events one user experienced. The death rate of moving events, top events, bottom events are also counted like top events. In order to classify the level of total death rate, the total death rates of all the users are ranked. Then the average of total death rates of all testers is counted and used as the threshold. The total death rates larger than threshold are high and smaller than threshold are low. Table 2 shows the categories of game events and values of death rate.

IV. METHOD

A. DATA PRE-PROCESSING

Before the experiments, data of questionnaires and game events are need to pre-process. The user characteristics and subjective evaluation in the questionnaire include 'Age', 'Gender', 'Parkour game experience', 'Physical exercise frequency', 'Character traits', 'Basic heartbeat', 'Diastolic pressure', 'Systolic pressure', 'Oxygen saturation', 'First game experience' and 'Willing to continue playing' are converted to digital values, as shown in Table 3. The total death rate and death rates of four categories game events are also digitized and shown in Table 5.

After pre-processing, user characteristics, total death rate and death rate of four categories are used as features. Then the improved RIPPER algorithm is used to find out the relations among positive-negative UX, user characteristics and categories of game events, represented by rule sets.

B. ALGORITHM

As elaborated in previous sections, finding the relations among positive-negative UX, user characteristics and categories of game events are the main purpose of this work. Rule learning algorithms have the capability to explore the relations between values of attributes and sophisticated

TABLE 3. Digital values of the questionnaire attributes.

Attributes	Value	Digitized
Age	11-15 to 46-50	0 to 7
Gender	male female	1 0
Parkour game experience	rich poor	1 0
Physical exercise frequency	frequent occasional	1 0
Character traits	extroverted introverted	1 0
Basic heartbeat	normal abnormal	1 0
Systolic pressure	normal abnormal	1 0
Diastolic pressure	normal abnormal	1 0
Oxygen saturation	normal abnormal	1 0
First experience of this game	yes no	1 0
Willing to continue playing	yes no	1 0

TABLE 4. Digital values of death rate on various game events.

Category of game event	Value	Digitized
Death rate of fixed game events	high low	1 0
Death rate of moving game events	high low	1 0
Death rate of top game events	high low	1 0
Death rate of bottom game events	high low	1 0

categories. Furthermore, the relations are represented by rule sets that are easy to be understood [34]. Among effective rule learning algorithms, the calculation cost of reduced error pruning (REP) algorithm is quite expensive. The calculation of incremental reduced error pruning (IREP) is faster than REP and the error rate is more competitive. However, one disadvantage of rule learning algorithms is that they often scale poorly with the sample size, particularly on noisy data [35]. Consequently, researchers propose a number of modifications based on IREP and resulted in RIPPER algorithm [36]. RIPPER has lower error rate and better performance than IREP.

Due to the limited scale of samples and sparse features in this paper, RIPPER algorithm is improved to adapt for missing samples with specific feature values. Since RIPPER is an adaption of IREP, we will briefly describe the basic function of IREP, RIPPER and elaborate the improved part of algorithms below.

The goal of IREP is to generate rule sets that cover as many samples of one category as possible and identify it from other categories. This category is positive category and other categories call negative category. Before generating rule sets, data set are randomly partitioned into three subsets. A training set occupies 60%, a pruning set occupies 20%, and a testing set occupies 20%. IREP generates a original rule related to positive category through the sequential coverage on training set. That is, one rule at a time. In the two-category, a "rule" is simply a conjunction of multiple features and a category. In IREP, rule is grown by Foil information gain:

$$F_{Gain} = m'_{+}(log_2 \frac{m'_{+}}{m'_{+} + m'_{-}} - log_2 \frac{m_{+}}{m_{+} + m_{-}}) \tag{1}$$

where m_+, m_- are the number of positive and negative samples covered by the old rule, and m'_+, m'_- are the number of positive and negative samples covered by the new rule after adding candidate feature value. Since the number of positive cases in the data is less than the negative cases, the Foil information gain firstly takes the positive cases covered by the new rules as the weights.

Rules begin with an empty conjunction of conditions, and consider adding to any condition with the form $A_n = v$. In this formula, A_n is a nominal attribute and v is a legal value for A_n. When a rule is growing, the category as the rule header. Any condition is traversed and added to the rule body as a candidate words. Rule body repeatedly adds condition that maximizes Foil's information gain criterion until the rule body only cover samples of positive category.

To prevent over fitting, the rule is immediately pruned to obtain the rule with high accuracy on pruning set after generating a rule on training set. After generating a new rule, IREP algorithm deletes any final sequence of conditions from the rule and chooses the deletion that maximizes function:

$$v(Rule, PrunePos, PruneNeg) \equiv \frac{p + (N - n)}{P + N} \tag{2}$$

where P (respectively N) is the total number of positive samples in pruning set (negative samples in pruning set), p(n) is the number of positive samples in pruning set (negative samples in pruning set) covered by rule. This process is repeated until no deletion improves the value of v.

After a rule is generated, the samples of positive category that have been covered by generated rule are removed, and the next rule is generated based on the remaining samples until the rule sets overwrite all positive samples in training set. IREP can automatically remove irrelevant factors after generating all rule sets and the factors related positive categories are leaved.

RIPPER makes modifications based on IREP and performs efficiently on large noisy data set. Using Equation (2) mentioned above, IREP may cause occasional incorrect pruning. For example, assuming that P and N are fixed, rule R_1 covers $p_1 = 1000$ positive samples and $n_1 = 500$ negative samples, rule R_2 covers $p_1 = 500$ positive samples and $n_1 = 1$ negative samples. The equation prefer R_1 to R_2. However, R_2 is highly accurate and R_1 is not. Based on this error, RIPPER replaced IREP's pruning equation with

$$v(Rule, PrunePos, PruneNeg) \equiv \frac{p - n}{p + n} \tag{3}$$

the meaning of p(n) in Equation (3) is the same as that of p(n) in Equation (2).

To reduce inaccuracies of rule sets caused by false pruning, RIPPER optimize the rule sets after all rule sets are generated by IREP. That is, RIPPER post-processes rules generated by IREP. For each rule, two variants are generated one by one in sequence, which are respectively introduced into the entire rule sets, and the optimal rule is reserved.

For each rule r_i in the rule set, two variants are made as follows:

r'_i: reusing $IREP^*$ algorithm to generate a rule r'_i based on the example covered by r_i, which is called a replacement rule.

r''_i: adding value of attributes to r_i, then use $IREP^*$ prune to generate a rule r''_i, this rule is called revision rule.

This paper improves Foil information gain in RIPPER. RIPPER will produce occasional failure if the data set miss samples with specific attribute values. There are two situations may arise:

1) If the values of an attribute are binary such as 1 and 0, and the samples covered by value 1 of this attribute are all negative samples; after old rule add the value 1 of this attribute, the rule traverse samples with the value 1 of this attribute in training set and calculates new Foil information gain; the new rule overlays some samples in training set, but the samples are all negative.
2) If the values of the attribute are similar to situation 1) and none of the samples has a value of 1 for this attribute; when the rule calculates new Foil information gain; the new rule does not cover any samples in the training set.

The first situation makes m'_+ in $log_2\frac{m'_+}{m'_+ + m'_-}$ be 0, the real number in log_2 is 0, and the result of $log_2\frac{m'_+}{m'_+ + m'_-}$ is negative infinity. In the second situation, the numerator and denominator of $\frac{m'_+}{m'_+ + m'_-}$ are both 0.

The improved algorithm is shown in Algorithm 1. For solving both of these cases, Foil information gain are improved with

$$F'_{Gain} = m'_+(log_2\frac{m'_+ + eps}{m'_+ + m'_- + 1} - log_2\frac{m_+}{m_+ + m_-}) \quad (4)$$

Based on the old Foil gain, eps is added to the numerator of log_2 with the value of 2^{-52}, which avoids 0 on the numerator and the true number of 0 in log_2, which in turn avoids negative occurrence of $log_2\frac{m'_+}{m'_+ + m'_-}$. The addition of 1 to the denominator of $\frac{m'_+}{m'_+ + m'_-}$ avoids the situation 2) where the $m'_+ + m'_-$ is 0 and the real number of log_2 in the situation 1) is a small negative value.

To objectively verify the rule sets related to UX, ECG signals of users are captured. Since each user's ECG baseline is different [9], the baseline of each user is adjusted before the experiments. The continuous RR interval during the game is measured by the equipment, and is used to obtain accurate heartbeat values. Based on the heartbeat values, Heart Rate Variability (HRV) at each moment of different user is calculated. These variability represents the physiological experience of users when they come across the defined game events, as shown in Figure 1.

Algorithm 1 The improved RIPPER algorithm.

Input:
GrowPos, GrowNeg, PrunePos, PruneNeg P_0,
Output:
Ruleset

```
Ruleset := null,Rule := null
repeat
    repeat
        Feature = ImprovFoil(GrowPos,GrowNeg)
        Rule:=GrowRule(Rule,Feature)
        Cov:=RuleCov(Rule,GrowPos)
    until Cov ∈ GrowPos
    Rule:=PruneRule(Rule,PrunePos,PruneNeg)
    add Rule to Ruleset
    remove samples covered by Rule from −(GrowPos,GrowNeg)
until GrowPos := null
```

V. EXPERIMENT AND RESULTS

A. TEST ENVIRONMENT

Due to limited resources, 300 testers are involved in this experiment. Before starting the game, a questionnaire (including 'Age', 'Gender', 'Parkour game experience', 'Physical exercise frequency', 'Character traits', 'Basic heartbeat', 'Diastolic pressure', 'Systolic pressure', 'Oxygen saturation', 'First game experience' and 'Willing to continue playing') marked with anonymous identity is assigned to each tester. In the questionnaire, the first five items are needed to fill by tester, 'basic heartbeat' means each user's ECG baseline, and the last two items are filled after playing.

In order to accurately count game events and match the game event with heart rate difference, the playing process of each user is recorded to get the game events and relevant times. As shown in Figure 2, continuous heartbeat signal of each tester during playing the game are obtained and each user's game events are recorded by camera. Each tester's game time is no less than five minutes so as to obtain sufficient data. Finally, testers have to answer whether he/she has the willing to continue playing on questionnaires.

B. TEST RESULTS

After testing, there are 275 testers left after removing testers with missing video or ECG data. Among them, the training set has 165 samples, the pruning set has 55 samples and the testing set has 55samples. All testers' ECG signals and HRV values in each second are pre-processed. Besides, to precisely calculate death rate and easily observe the HRV values of different categories of game events, we draw complete heart rate variability graph of each tester and manually marked all events defined in Section III-B on the graph. Figure 1 shows

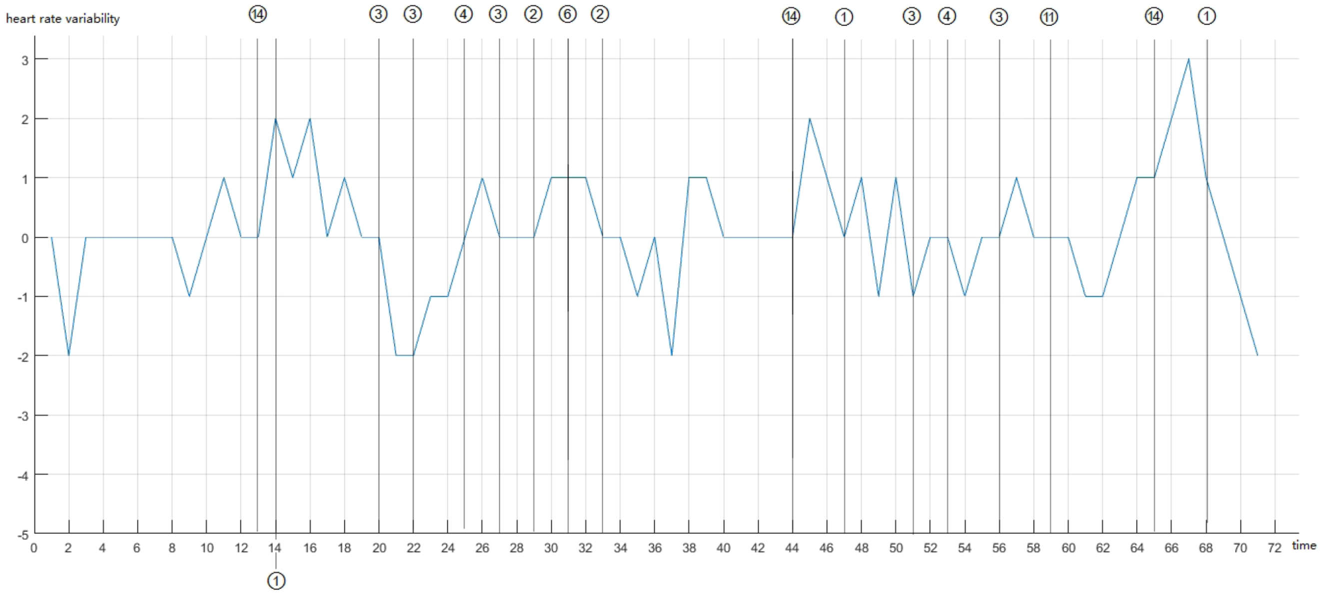
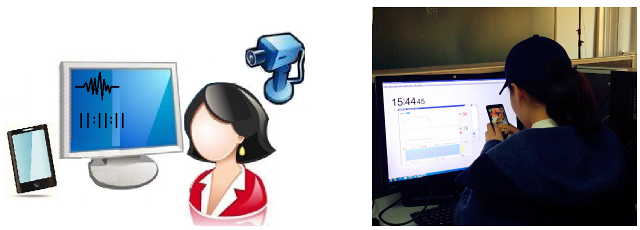


FIGURE 1. Events marked on the HRV chart. (1 Iron chain over the cliff. 2 Poll on the ground. 3 Cliff. 4 Stone on the ground. 6 Red brick on the ground. 11.Fixed horizontal trees over the ground. 14 Running bear behind to catch the player.).



(a) (b)

FIGURE 2. Experiment environment. (a) Environment configuration. (b) Real test environment.

an example of the game events on the HRV graph. According to HRV graph of each tester, total death rate and death rates of the four categories game events of each tester are counted and digitized according to the method in Section IV-A.

After pre-processing, the relations among user characteristics, categories of game events and positive-negative UX are learned by the improved RIPPER algorithm. Since the number of samples in this experiment is small, this paper uses five-fold cross-validation. As a result, five kinds of data sets are obtained. The accuracy of positive and negative rule sets in different training groups are shown in Table 5. The results show that the positive rule sets learned on the third training sets and pruning sets are more precise, and the negative rule sets learned on the second training sets and pruning sets are more precise.

The rule sets positively related to UX are listed below:

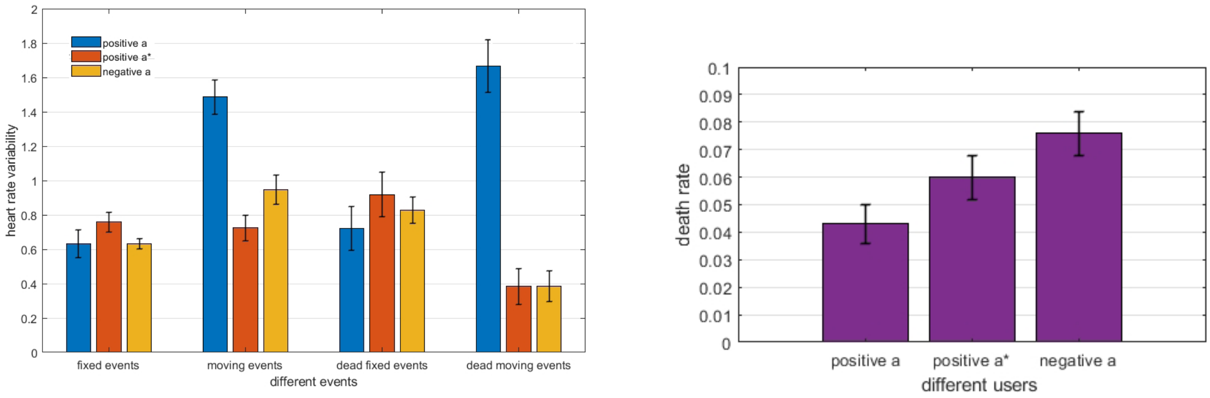
$$(a)\begin{cases}(positive\ user\ experience) \leftarrow (16-20)) \cap \\ (low\ death\ rate\ of\ fixed\ game\ events) \cap \\ (low\ total\ death\ rate)\end{cases}$$

TABLE 5. The accuracy of positive and negative rule sets in different training group.

Training group	Positive rules Accuracy	Negative rules Accuracy
Group1	50%	66.7%
Group2	66.7%	**85.7%**
Group3	**98%**	83.3%
Group4	36.4%	75%
Group5	75%	77.8%

Rule(a) relates to low total death rate and low death rate of fixed game events. To objectively verify rule(a), the average total death rate, and the average HRV values of fixed events, moving events, dead fixed events and dead moving events of 'positive a', 'positive a^*' and 'negative a' are counted respectively. To illustrate the reliability of these average HRV values, confidence interval is used to estimate the range of average total deth rate and the average heartbeat range of the game events mentioned above. As Figure 3(a) shows, the average HRV value of 'positive a' rises significantly but the values of 'positive a^*' and 'negative a' decline when compared fixed events with moving events and compared dead fixed events with dead moving events. In Figure 3(b), the death rate of 'positive a' is slightly smaller than 'positive a^*' and 'negative a'. With the conclusion, the positive UX of young users with age from 16-20 are sensitive to the configuration of fixed events and overall difficulty. Low difficulty or fewer fixed events combine low overall difficulty attracts such person.

$$(b)\begin{cases}(positive\ user\ experience) \leftarrow (male) \cap \\ (extroverted) \cap (first\ experience\ of\ this\ game) \cap \\ (low\ death\ rate\ of\ \ bottom\ game\ events)\end{cases}$$



(a)

(b)

FIGURE 3. **Rule(a)-1 shows the average HRV values of different users in different events relate to rule(a), 'positive a' means positive samples covered by rule(a), 'positive a^*' means positive samples not covered by rule(a), 'negative a' means negative samples not covered by rule(a). Rule(a)-2 shows total death rate of different user relate to rule(a). (a) rule(a)-1. (b) rule(a)-2.**

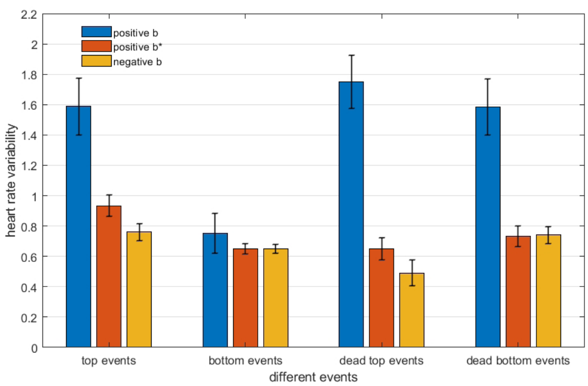


FIGURE 4. **This figure shows the average HRV values of different users in different events relate to rule(b). 'positive b' means positive samples covered by rule(b). 'positive b^*' means positive samples not covered by rule(b). 'negative b' means negative samples not covered by rule(b).**

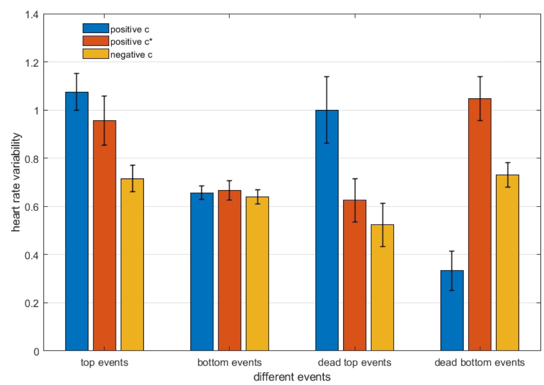


FIGURE 5. **This figure shows the average HRV values of different users in different events relate to rule(c). 'positive c' means positive samples covered by rule(c). 'positive c^*' means positive samples not covered by rule(c). 'negative c' means negative samples not covered by rule(c).**

Rule(b) relates to low death rate of bottom game events. Similarly, the average HRV values of top events, bottom events, dead top events and dead bottom events of 'positive b', 'positive b^*' and 'negative b' are counted respectively. As Figure 4 shows, the average HRV value of 'positive b' declines but the values of 'positive b^*' and 'negative b' rise when compared dead top events with dead bottom events. Therefore, the positive UX of the extroverted men who first experience this game are related to the configuration of bottom events. Low difficulty bottom events attracts such person.

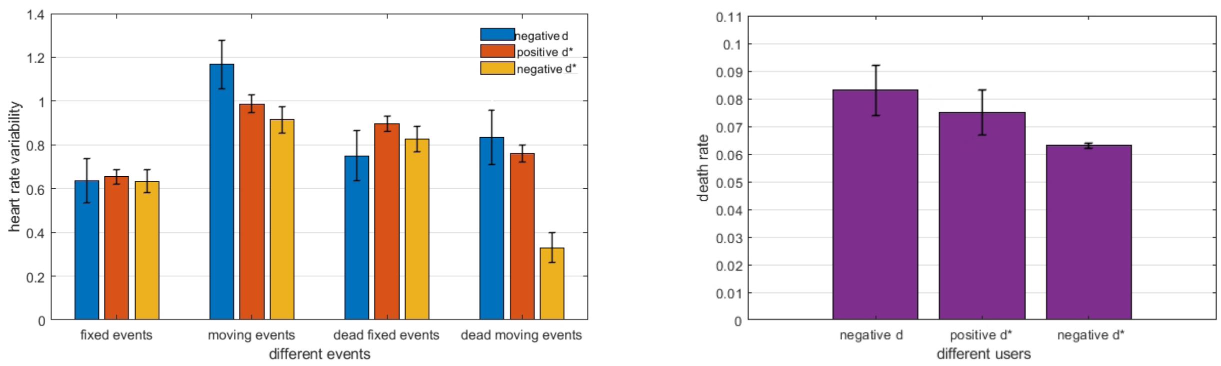
$$(c)\begin{cases}(positive\ user\ experience) \leftarrow (21-25) \cap \\ (introverted) \cap (less\ physical\ exercise\ frequency) \cap \\ (low\ death\ rate\ of\ bottom\ game\ events)\end{cases}$$

Rule(c) relates to low death rate of bottom events. The average HRV values of different events are counted like rule(b). As Figure 5 shows, the average HRV value of 'positive c' declines but the values of 'positive c^*' and 'negative c' rises when compared dead top events with dead bottom events. The the positive UX of introverted young users with age of 21-25 and less physical exercise frequency are sensitive to the configuration of bottom events. Low difficulty bottom events attracts such person.

The rule sets negatively related to UX are listed below:

$$(d)\begin{cases}(negative\ user\ experience) \leftarrow (male) \cap \\ (introverted) \cap (high\ total\ death\ rate) \cap \\ (low\ death\ rate\ of\ fixed\ game\ events)\end{cases}$$

Rule(d) relates to low death rate of fixed events. As Figure 6(a) shows, the average HRV value of 'negative d' rises but the values of 'positive d^*' and 'negative d^*' decline when compared dead fixed events with dead moving events. Rule(d) also relates to high total death rate, the death rate of 'negative d' is slightly larger than 'positive d^*' and 'negative d^*' in Figure 6(b). The negative UX of the introverted men



(a) (rule(d)-1) (b) rule(d)-2

FIGURE 6. Rule(d)-1 shows the average HRV values of different users in different events relate to rule(d), 'negative d' means negative samples covered by rule(d), 'positive $d*$' means positive samples not covered by rule(d), 'negative $d*$' means negative samples not covered by rule(d). Rule(d)-2 shows total death rate of different users relate to rule(d). (a) (rule(d)-1). (b) rule(d)-2.

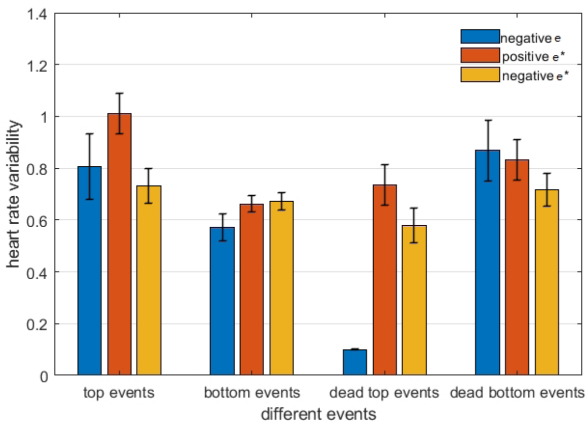


FIGURE 7. This figure shows the average HRV values of different users in different events relate to rule(e). 'negative e' means negative samples covered by rule(e). 'positive e*' means positive samples not covered by rule(e). 'negative e*' means negative samples not covered by rule(e).

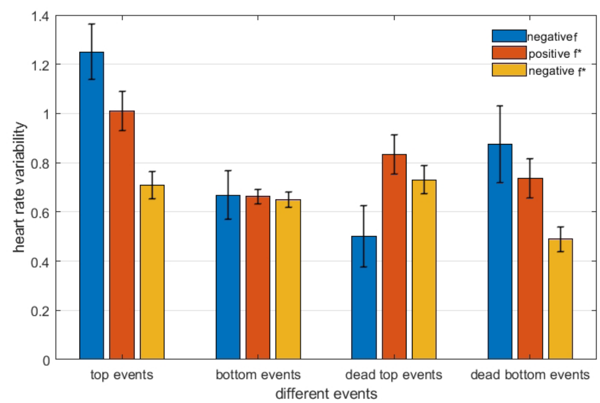


FIGURE 8. This figure shows the average HRV values of different users in different events relate to rule(f). 'negative f' means negative samples covered by rule(f). 'positive $f*$' means positive samples not covered by rule(f). 'negative $f*$' means negative samples not covered by rule(f).

are related to the configuration of fixed events and overall difficulty. Low difficulty fixed events combine high overall difficulty attract such person.

$$(e)\begin{cases}(negative\ user\ experience) \leftarrow (male) \cap \\ (rich\ parkour\ game\ experience) \cap \\ (low\ death\ rate\ of\ top\ game\ events)\end{cases}$$

Rule(e) relates to low death rate of top events. As Figure 7 shows, the average HRV values of 'negative e' rises significantly but the values of 'positive e^*' and 'negative e^*' rise a little bit when compared dead top events with dead bottom events. The negative UX of the men who have rich parkour game experience are more sensitive to the configuration of top events. Low difficulty top events attracts such person.

$$(f)\begin{cases}(negative\ user\ experience) \leftarrow (introverted) \cap \\ (rich\ parkour\ game\ experience) \cap \\ (low\ death\ rate\ of\ bottom\ game\ events) \cap \\ (low\ physical\ exercise\ frequency)\end{cases}$$

Rule(f) relates to high death rate of bottom events. As Figure 8 shows, the average HRV value of 'negative f' rises but the values of 'positive f^*' and 'negative f^*' decline when compared dead top events with dead bottom events. The configuration of bottom events would affect the negative UX of the introverted users who have rich parkour game experience and low physical exercise frequency. Low difficulty bottom events attracts such person.

$$(g)\begin{cases}(negative\ user\ experience) \leftarrow (16-20) \cap \\ (not\ first\ experience\ of\ this\ game) \cap \\ (high\ total\ death\ rate)\end{cases}$$

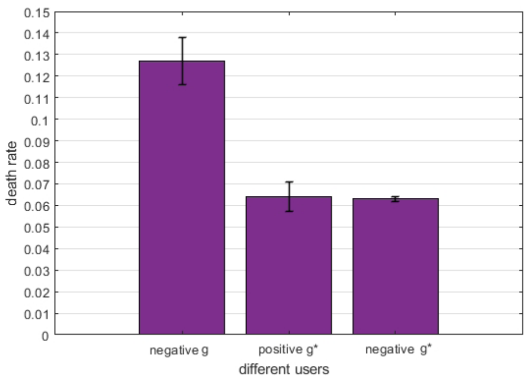


FIGURE 9. This figure shows average total death rate of different user relate to rule(g). 'negative g' means negative samples covered by rule(g). 'positive g^*' means positive samples not covered by rule(g). 'negative g^*' mean negative samples not covered by rule(g).

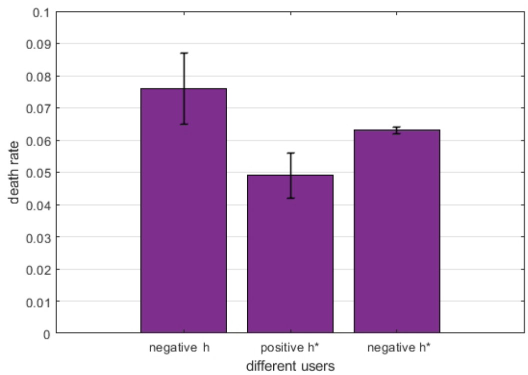


FIGURE 10. This figure shows average total death rate of different user relate to rule(h). 'negative h' means negative samples covered by rule(h). 'positive h^*' means positive samples not covered by rule(h). 'negative h^*' means negative samples not covered by rule(h).

As Figure 9 show that average total death rate of 'negative g' obviously higher than 'positive g^*' and 'negative g^*'. The poor configuration of overall difficulty would bring more negative UX of the 16-20 years old users who has played this game.

$$(h)\begin{cases}(negative\ user\ experience) \leftarrow \cap\,(extroverted)\cap\\(poor\ parkor\ game\ experience)\cap\\(not\ first\ experience\ of\ this\ game)\cap\\(high\ total\ death\ rate)\end{cases}$$

Similarly, as Figure 10 shows that average total death rate of 'negative h' higher than 'positive h^*' and 'negative h^*'. The negative UX of extroverted users that have poor parkour game experience and have played this game are sensitive to the configuration of overall difficulty. Low overall difficulty attracts such person.

VI. CONCLUSION AND FUTURE WORK

The goal of this paper is to find user characteristics and categories of game events that are positively or negatively related to UX. To obtain sufficient factors, this paper reviews a large number of literature and then designs a questionnaire includes comprehensive user characteristics related to UX combined with subjective evaluation. Game events are defined according to operation and events of game. Before playing, 300 testers are asked to fill questionnaire and relax to obtain each user's ECG baseline. Then they play 'Temple Run' with professional ECG instrument at least 5 minutes. The playing progresses are recorded by camera to obtain time of game events. The ECG signals are used to verify UX objectively. After playing, user characteristics, game events, game time and ECG data are collected, digitized and analyzed. HRV are calculated by ECG baseline at resting stage and continuous ECG during game.

After pre-processing data, 275 samples are left and the number of samples with abnormal basic heart rate or systolic pressure or diastolic pressure or oxygen saturation is 0, but which does not mean that these four attributes are not related to UX. In this experiment, game events are marked on the digitized HRV graph, which can accurately calculate each tester's death rates of different categories events. The RIPPER algorithm are used to learn the relations among user characteristics, categories of game events and positive-negative UX. User characteristics include 'gender', 'age', 'Parkour experience', 'physical exercise frequency', 'character traits', 'basic heartbeat', 'systolic presure', 'diastolic pressure', 'oxygen saturation', 'first experience of this game'. Categories of game events include 'total events', 'top events', 'bottom events', 'fixed events' and 'moving events'. The Foil information gain in the RIPPER algorithm are improved by this paper since missing samples with specific attribute values. To fully prove the accuracy of the rule sets generated by RIPPER, the average HRV values of game events covered by rule sets are counted respectively. The experimental results show that the accuracy of rule sets is 98%, and the rule sets are objectively proved by the average HRV values of different categories events.

This experiment only tests 300 testers and large-scale testers will be tested in the future. In order to make the conclusion more general, the age range for collecting samples will be also expanded in the future experiments. Since the tester's characteristics in the questionnaire is only binarized, values of attributes will be more subdivided in further researches. Additionally, more relevant characteristics will be explored and added in our future works, such as user work, pleasure motivation, game price, etc.

REFERENCES

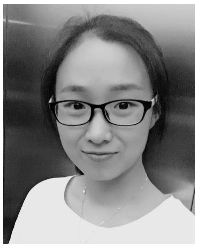
[1] Y. Wang and X. Che, "How to keep people playing mobile games: An experience requirements testing approach," in *Proc. Int. IEEE Conf. Ubiquitous Intell. Comput., Adv. Trusted Comput., Scalable Comput. Commun., Cloud Big Data Comput., Internet People, Smart World Congr. (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld)*, Toulouse, France, Jul. 2016, pp. 815–822.

[2] I. O. Pappas, P. Mikalef, P. Kourouthanassis, and M. N. Giannakos, "Fuzzy-set analysis to understand user experience in mobile applications," in *Proc. 23rd Amer. Conf. Inf. Syst. (AMCIS)*, Boston, MA, USA, Aug. 2017, pp. 112–119.

[3] A. Perttula, P. Tuomi, K. Kiili, M. Suominen, A. Koivisto, and J. Multisilta, "Enriching shared experience by collective heart rate," *Int. J. Social Humanistic Comput.*, vol. 2, nos. 1–2, pp. 31–50, 2013.
[4] Z. Ma, X. Chang, Z. Xu, N. Sebe, and A. G. Hauptmann, "Joint attributes and event analysis for multimedia event detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 7, pp. 2921–2930, Jul. 2018.
[5] M. Huang, R. Ali, and J. Liao, "The effect of user experience in online games on word of mouth: A pleasure-arousal-dominance (PAD) model perspective," *Comput. Hum. Behav.*, vol. 75, pp. 329–338, Oct. 2017.
[6] J. Harms, D. Seitz, C. Wimmer, K. Kappel, and T. Grechenig, "Low-cost gamification of online surveys: Improving the user experience through achievement badges," in *Proc. Annu. Symp. Comput.-Hum. Interact. Play (CHI PLAY)*, London, U.K., Oct. 2015, pp. 109–113.
[7] M. Luo, X. Chang, Z. Li, L. Nie, A. G. Hauptmann, and Q. Zheng, "Simple to complex cross-modal learning to rank," *Comput. Vis. Image Understand.*, vol. 163, pp. 67–77, Oct. 2017.
[8] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 2, pp. 156–166, Jun. 2005.
[9] E. P. N. Castellar, K. Oksanen, and J. Van Looy, "Assessing game experience: Heart rate variability, in-game behavior and self-report measures," in *Proc. 6th Int. Workshop Qual. Multimedia Exper. (QoMEX)*, Singapore, Sep. 2014, pp. 292–296.
[10] M. Schrepp, M. P. Cota, R. Gonçalves, A. Hinderks, and J. Thomaschewski, "Adaption of user experience questionnaires for different user groups," *Universal Access Inf. Soc.*, vol. 16, no. 3, pp. 629–640, 2017.
[11] S. Wang, X. Chang, X. Li, G. Long, L. Yao, and Q. Z. Sheng, "Diagnosis code assignment using sparsity-based disease correlation embedding," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 12, pp. 3191–3202, Dec. 2016.
[12] L. Yao *et al.*, "Using physiological measures to evaluate user experience of mobile applications," in *Proc. 11th Int. Conf. Eng. Psychol. Cogn. Ergonom. (EPCE)*, Heraklion, Greece, Jun. 2014, pp. 301–310.
[13] S. Wang, X. Li, X. Chang, L. Yao, Q. Z. Sheng, and G. Long, "Learning multiple diagnosis codes for ICU patients with local disease correlation mining," *ACM Trans. Knowl. Discovery Data*, vol. 11, no. 3, 2017, Art. no. 31.
[14] K. Kiili, A. Perttula, S. Arnab, and M. Suominen, "Flow experience as a quality measure in evaluating physically activating serious games," in *Proc. 2nd Int. Conf. Games Learn. Alliance (GALA)*, Paris, France, Oct. 2013, pp. 200–212.
[15] M. Schrepp, A. Hinderks, and J. Thomaschewski, "Applying the user experience questionnaire (UEQ) in different evaluation scenarios," in *Proc. 3rd Int. Conf. Design, User Exper. Usability (DUXU)*, Heraklion, Greece, Jun. 2014, pp. 383–392.
[16] K. Tcha-Tokey, E. Loup-Escande, O. Christmann, and S. Richir, "A questionnaire to measure the user experience in immersive virtual environments," in *Proc. Virtual Reality Int. Conf.*, Laval, France, Mar. 2016, pp. 19-1–19-5.
[17] S. Zhang, S. Zhao, Y. Sui, and L. Zhang, "Single object tracking with fuzzy least squares support vector machine," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5723–5738, Dec. 2015.
[18] S. Pan, J. Wu, X. Zhu, C. Zhang, and S. Y. Philip, "Joint structure feature exploration and regularization for multi-task graph classification," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 3, pp. 715–728, May 2016.
[19] R. Bernhaupt and M. Pirker, "Evaluating user experience for interactive television: Towards the development of a domain-specific user experience questionnaire," in *Proc. IFIP Conf. Hum.-Comput. Interact. (INTERACT)*, Cape Town, South Africa, Sep. 2013, pp. 642–659.
[20] M. I. M. Eid and H. I. Abbas, "User adaptation and ERP benefits: Moderation analysis of user experience with ERP," *Kybernetes*, vol. 46, no. 3, pp. 530–549, 2017.
[21] S. Kang, L. Norooz, V. Oguamanam, A. C. Plane, T. L. Clegg, and J. E. Froehlich, "SharedPhys: Live physiological sensing, whole-body interaction, and large-screen visualizations to support shared inquiry experiences," in *Proc. 15th Int. Conf. Interact. Design Children (IDC)*, Manchester, U.K., Jun. 2016, pp. 275–287.
[22] F. Wang, X. Chen, G. Sun, and X. Liang, "Study on the influence of drivers' physiological characteristics of urban bus stop," in *Proc. 19th Int. Conf. Hum.-Comput. Interact. (HCI)*, Vancouver, BC, Canada, Jul. 2017, pp. 263–272.
[23] A. M. Khan and M. Lawo, "Recognizing emotional states using physiological devices," in *Proc. 18th Int. Conf. Hum.-Comput. Interact. (HCI)*, Toronto, ON, Canada, Jul. 2016, pp. 164–171.
[24] N. Martin and J.-M. Diverrez, "From physiological measures to an automatic recognition system of stress," in *Proc. 18th Int. Conf. Hum.-Comput. Interact. (HCI)*, Toronto, ON, Canada, Jul. 2016, pp. 172–176.
[25] S. Zhang, Y. Sui, X. Yu, S. Zhao, and L. Zhang, "Hybrid support vector machines for robust object tracking," *Pattern Recognit.*, vol. 48, no. 8, pp. 2474–2488, 2015.
[26] S. Pan, J. Wu, X. Zhu, G. Long, and C. Zhang, "Finding the best not the most: Regularized loss minimization subgraph selection for graph classification," *Pattern Recognit.*, vol. 48, no. 11, pp. 3783–3796, 2015.
[27] S. Pan, J. Wu, X. Zhu, G. Long, and C. Zhang, "Task sensitive feature exploration and learning for multitask graph classification," *IEEE Trans. Cybern.*, vol. 47, no. 3, pp. 744–758, Mar. 2017.
[28] M. S. Perez-Rosero, B. Rezaei, M. Akçakaya, and S. Ostadabbas, "Decoding emotional experiences through physiological signal processing," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, New Orleans, LA, USA, Mar. 2017, pp. 881–885.
[29] R. L. Mandryk, K. M. Inkpen, and T. W. Calvert, "Using psychophysiological techniques to measure user experience with entertainment technologies," *Behav. Inf. Technol.*, vol. 25, no. 2, pp. 141–158, 2006.
[30] L. Alexandros and X. Michalis, "The physiological measurements as a critical indicator in users' experience evaluation," in *Proc. 17th Panhellenic Conf. Inform. (PCI)*, Thessaloniki, Greece, Sep. 2013, pp. 258–263.
[31] I. Carrillo, V. Meza-Kubo, A. L. Morán, G. Galindo, and E. García-Canseco, "Emotions identification to measure user experience using brain biometric signals," in *Proc. Int. Conf. Hum. Aspects IT Aged Population (ITAP)*, Los Angeles, CA, USA, Aug. 2015, pp. 15–25.
[32] T. Lin, A. Maejima, and S. Morishima, "Using subjective and physiological measures to evaluate audience-participating movie experience," in *Proc. Work. Conf. Adv. Vis. Interfaces (AVI)*, Napoli, Italy, May 2008, pp. 49–56.
[33] G. Gonçalves, E. Mourão, L. Torok, D. Trevisan, E. Clua, and A. A. Montenegro, "Understanding user experience with game controllers: A case study with an adaptive smart controller and a traditional gamepad," in *Proc. 16th Int. Conf. Entertainment Comput. (ICEC)*, Tsukuba, Japan, Sep. 2017, pp. 59–71.
[34] J. Catlett, "Megainduction: A test flight," in *Proc. 8th Int. Workshop (ML)*. Evanston, IL, USA: Northwestern Univ., 1991, pp. 596–599.
[35] W. W. Cohen, "Efficient pruning methods for separate-and-conquer rule learning systems," in *Proc. 13th Int. Joint Conf. Artif. Intell.*, Chambéry, France, Aug./Sep. 1993, pp. 988–994.
[36] W. W. Cohen, "Fast effective rule induction," in *Proc. 12th Int. Conf. Mach. Learn.*, Tahoe City, CA, USA, Jul. 1995, pp. 115–123.

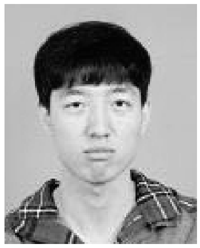
QI YU received the bachelor's degree in software engineering from the School of Software, North University of China, Taiyuan, China, in 2016. She is currently pursuing the master's degree with the School of Software Engineering, Beijing Jiaotong University. Her research interests include data analysis, machine learning, testing, and user experience.

XIAOPING CHE (M'16) received the B.S. degree in network engineering from the Beijing University of Posts and Telecommunications, Beijing, China, in 2009, the M.S. degree in computer and communication networks from Telecom SudParis, France, in 2011, and the Ph.D. degree from the Institute Mines-Telecom/Telecom SudParis, France, in 2014. From 2011 to 2014, he was with the French CNRS laboratory SAMOVAR. He is currently an Associate Professor with the School of Software Engineering, Beijing Jiaotong University. His main research interests include user experience, passive testing, crowdsensing, and Internet of Things. He is a member of the ACM.



SIQI MA received the bachelor's degree in software engineering from the School of Software Enginerring, Beijing Jiaotong University, Beijing, China, in 2017, where she is currently pursuing the master's degree. In 2016, she has been to EPITA for one semester as an Exchange Student. Her research interests include data analysis, machine learning, testing, and user experience.

SHIRUI PAN received the Ph.D. degree in computer science from the University of Technology Sydney, NSW, Australia, in 2015. He is currently a Lecturer with the Centre for Artificial Intelligence, Faculty of Engineering and Information Technology, University of Technology Sydney. He has published over 40 research papers in top-tier journals and conferences, including the IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, the IEEE TRANSACTIONS ON CYBERNETICS, the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, *Pattern Recognition*, IJCAI, ICDE, ICDM, SDM, CIKM, and PAKDD. He serves as an Associate Editor of the IEEE ACCESS. His current research interests include data mining and machine learning.



YUXIANG YANG received the bachelor's degree in computer science and technology from the Northeastern University of China, Liaoning, China, in 2014. He is currently pursuing the Ph.D. degree with the School of Software Engineering, Beijing Jiaotong University. His research interests include image processing, deep learning, reinforcement learning, and object tracking.

WEIWEI XING received the B.S. degree in computer science and technology and the Ph.D. degree in signal and information processing from Beijing Jiaotong University, Beijing, China, in 2001 and 2006, respectively. She is currently a Professor with the School of Software Engineering, Beijing Jiaotong University. Her research interests include intelligent information processing and machine learning.

XIMENG WANG received the B.E. degree in communication engineering from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2011, and the M.E. degree in software engineering from Beijing Jiaotong University, Beijing, China, in 2013. He is currently pursuing the dual Ph.D. degrees with the Beijing Jiaotong University and the University of Technology Sydney. His current research interests include recommender systems, network embedding, and data mining.