

Affective Classification of Gaming Activities Coming From RPG Gaming Sessions

Fabrizio Balducci, Costantino Grana

Dipartimento di Ingegneria “Enzo Ferrari”
Università degli Studi di Modena e Reggio Emilia
Via Vivarelli 10, Modena MO 41125, Italy
{name.surname}@unimore.it

Abstract. Each human activity involves feelings and subjective emotions: different people will perform and sense the same task with different outcomes and experience; to understand this experience, concepts like Flow or Boredom must be investigated using objective data provided by methods like electroencephalography. This work carries on the analysis of EEG data coming from brain-computer interface and videogame “Neverwinter Nights 2”: we propose an experimental methodology comparing results coming from different off-the-shelf machine learning techniques, employed on the gaming activities, to check if each affective state corresponds to the hypothesis fixed in their formal design guidelines.

Keywords: classification, EEG, brain-computer interfaces, games

1 Introduction

The concepts of ‘play’ and ‘learn’ are strictly connected at various degrees: the best method to learn how to safely execute a task is to transform an assignment in a playing activity from which to gain experience for future goals [7]: this is the basis of the *gamification* and *serious gaming* theories which try to introduce playful aspects in all the daily interactive tasks. Lindley [4] defines a game as a goal-directed and competitive activity conducted with agreed rules: “to play” involves learning the internal rules and mechanics that characterize its dynamics and denote each game genre (the gameplay); it becomes crucial to consider player’s feelings and preferences to personalize the player experience tailoring each gaming session to specific needs and desires.

Studies from Mandryk *et al.* [11] have successfully demonstrated how psychophysiological techniques (like EEG) evidence human emotions and cognitive activity. The terms “affect” and “emotion” are often used interchangeably and refer to a short-time emotional peak while a “mood” or “affective state” denotes a continuous lasting emotional trend which may involve more emotions and may influence the reactions [6]. Brainwaves allow to evaluate basic emotions and infer the affective states that a subject experiences. A useful scheme to evaluate basic emotions is the *Circumplex model of affect* by Russell *et al.* [14]. In the

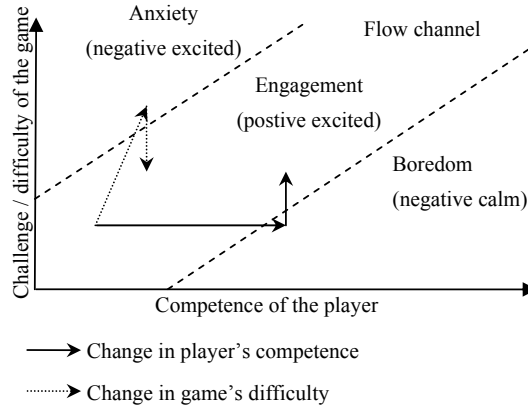


Fig. 1. The flow zone

field of computer science BCI has been used with various purposes like measure mnemonic and cognitive efforts [8], classify tasks [10] and improve usability [15]. In this paper we use different machine learning approaches to infer, from organized EEG data, the affective state (namely Boredom or Flow) experienced by the player: different classifiers are compared and tested on two game levels designed to induce specific sets of emotions. In this study we use a virtual world as research environments because it is cheap and easy to monitor and control, but it is easy to extend the proposed methodology to real-world (learning) tasks, for example by designing different time-limited versions of them and integrate the classic evaluation methodologies (interviews, think aloud, cognitive walkthrough) with objective evaluation coming from physiological data.

Section 2 summarizes the two proposed game levels and Section 3 briefly illustrates the technical system architecture, the experimental setup and the data format. Section 4 describes the proposed methodology for data analysis and Section 5 shows the classification results from three machine learning classifiers; finally, conclusions and suggestions for future work are drawn in Section 6.

2 Affective design for Role-Playing videogames

As shown in Lankoski [9] the RPG genre is based on statistics, object inventory and environmental exploration and its gameplay focuses on strong story plot, choices, skill progression, proactive and interactive allies, world exploration, people collaboration, interaction based on dialogues.

The works of Balducci *et al.* [1, 2] takes inspiration from what Nacke *et al.* [12] has experienced with First-Person Shooter (FPS) genre: two sets of formal design guidelines, with one of their possible development, were proposed for a RPG game with the aim to induce the Boredom affective state and the Flow one.

Csikszentmihályi [5] defines *Boredom* like a state in which player's skills are greater than required: the proposed game level features linearity and repeti-



Fig. 2. The same activity (dialogue) differently developed for the two game levels



Fig. 3. Activities in the flow game level: “chest opened” and “group fight” with allies

tiveness with poor gratification, minimal plot-story, weak visual assets, plain dialogues (Fig. 2) and unnecessary allies.

The notion of *Flow* is characterized by constant balance between challenges and skills: hard challenges will produce *anxiety* while very high skills will increase the perceived boredom (Chanel *et al.* [3], Fig. 1). The proposed game level features complex dialogues and multiple goals, proactive and interactive allies which helps to accomplish profitable activities while the level structure and the visual assets encourage environmental exploration (Fig. 3).

3 Experimental setup

The Emotiv EPOC headset is a wireless neuro-signal system with 14 wet sensors (+2 reference) capable of detecting brainwaves at 128 Hz sequential sampling rate; sensors are placed around the pre-frontal and frontal brain regions according to the international 10-20 standard system. The headset exposes two types of data: the raw (from each sensor) and the pre-classified by internal algorithms (Engagement, Excitement, Frustration, Meditation, Long-Term Excitement); to characterize the gaming activities we will consider the second set.

Neverwinter Nights 2 has a visual level editor with a scripting language that allows to manage internal game variables and permits to customize the User Interface. We used the NWNX [16] tool to deploy a C++ plugin that permits the communication between the game and the headset allowing textual messages exchange in the server version of the game; in this way, a game script calls

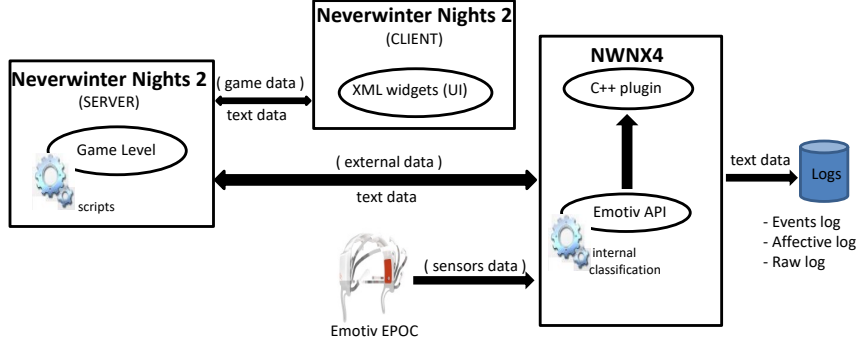


Fig. 4. The architecture for the data gathering tool

the extern plugin to iteratively manage the internal (game variables) and the external (EEG) data. The scheme of the system architecture is in Fig. 4.

The experiment has a two-treatment (boredom/flow) within-subject design, with ‘game level’ as independent variable factor; the dependent variables are the five pre-classified emotion summarized and controlled by our method.

The subjects involved are 19 students (13 male, 6 female) which play both game levels in each session. The *Setup* step consists in headset and sensors placing with signal tuning and, after this, a brief explanation about the study is provided followed by a pre-questionnaire about subject’s gaming preferences; next the *Tutorial* game level allows to familiarize with the commands and the user-interface (we also consider this steps a way to induce relaxation and a neutral initial affective mood, similarly to [17, 13]). Each subject plays with the same game character with all RPG statistics set to average values and, at the end of a session, a post-questionnaire about the player experience is compiled.

After each experiment we have different data sources: a video capture of the gaming session, the pre- and post- questionnaires and three textual files featuring the synchronized game and EEG data.

The *events log* contains at each row the timestamp and the tag of the occurred event; an entry in *affective log* contains the same timestamp, the game area in which the player was (XY coordinates) and the punctual values of the five internal pre-classified emotions (values in range 0-1) while each row of the *raw log*, similar to the previous, collects the values of the 14 sensors.

4 Proposed methodology

In order to interpret data which greatly change over time, analyzing a gaming session using all the log entries results impractical due to the time length; furthermore each game activity can have different duration and it is hard to identify a precise moment in which an affective stimulus appears.

Our approach is to split the log, identifying limited sequences for each activity faced during the gaming session: we remove the rows that don’t belong to the

tasks and events designed with the formal guidelines and, in this way, we can isolate those semantic units of time and make them more easily treatable.

The boredom game level has an average of 14 activities while the flow one offers the double having more activity types (which are *simple/dubbed/riddle dialogue*, *single/group fight*, *chest open*, *skills upgrade* and *stealing task*); notice that the second game level results longer since reaching a flow state requires more time to evolve the RPG gameplay in a natural way to the player’s eyes.

We obtain our data by crossing the *events log* with the *affective log*: the first one provides the identifier and the initial time of each activity, while the second provides the affective values, synchronized by the same timestamp.

The fact that each game level has been designed and developed with formal guidelines permits to label as “Flow” or “Boredom” each of their activities depending on the membership: in this way we have a dataset implicitly annotated and so, using binary classification algorithms, it is possible to check if the affective data of an activity are characteristic enough to differentiate among the two levels; if this occurs for a substantial number of activities, then player’s emotions have been well manipulated between the gaming sessions.

To describe the affective mood of an activity and use it in a machine-learning algorithm, we take the five pre-classified EEG emotions and, for each of them, we compute five numerical *features*:

- the *angular coefficient* β and the *intercept* α from a regression line, calculated considering the time T as a positive variable which constantly increases on the x-axis; it suggests the presence of an affective trend for the emotion E
- the *Pearson correlation coefficient* between time T and an emotion E : for a sample of n rows, if $r_e > \pm 0.7$ then there is evidence of strong local correlation, direct (positive sign) or inverse (negative sign).
- the *arithmetic mean* \bar{E} and the *variance* σ_E^2 that give quantitative information on the affective values of emotion E .

This results in a 790x25 predictors matrix M : rows represent the labeled activities (277 boredom, 513 flow) and columns the features (five for each emotion); we balance this training dataset increasing the boredom cases by randomly sampling 236 examples from the original ones, finally giving a 1026x25 matrix.

5 Classification and Results

Three supervised classification techniques have been tested, trained with a “leave-one-out” setup: two employ Support Vector Machines (linear and Radial Basis Function kernel) and the third consists in binary Decision Trees. Results are provided in Table 1, Table 2 and Table 3 respectively.

The *Decision Trees* method has the best performance: it is able to correctly separate 96% of the gaming activities, confirming the correctness of design and development phases by following the guidelines and, moreover, the validity of the proposed methodology.

Table 1. SVM-linear classification results

	boredom	flow	Tot.
original events	513 (ext.)	513	1026
classified events	610	416	1026
true positives	408	311	719
false negatives	105	202	307
accuracy			0.73%

Table 2. SVM-RBF classification results

	boredom	flow	Tot.
original events	513(ext.)	513	1026
classified events	542	484	1026
true positives	489	460	949
false negatives	24	53	77
accuracy			0.92%

Table 3. Decision Trees classification results

	boredom	flow	Tot.
original events	513 (ext.)	513	1026
classified events	507	519	1026
true positives	488	494	982
false negatives	25	19	44
recall	0.95%	0.96%	
f1-score	0.96%	0.96%	
accuracy			0.96%

5.1 Classification for game level areas

Considering the areas by which each game level is composed, from Table 4 and 5 we can understand the best-designed parts: the 90% good classification for boredom areas suggests that they were constantly recognized as repetitive and tedious; it is also remarkable that while the first two areas have the same structure, the third reduces the perceived boredom introducing very small variations.

For the flow game level, the classification performance varies, but it is always above 95%: the correlation grows from 95.4% for the first area up to 98.3% for the fourth one; the second area contains the largest number of activities (163) so we observe a slight decrease with respect to the typical progressive involvement associated to the flow affective state (the fifth smallest area is not significant since it was reached by only 3 subjects).

5.2 Classification for activity types

Table 6 and Table 7 show which activity types are the best classified: in the boredom game level *chest opened* is the best classified (96.7%) followed by *dialogue* (95.5%) while all classification ratings are over 93%.

In the flow game level, with the activities specifically developed for its gameplay, *stealing action*, *skills upgrade* and *dubbed dialogue* are 100% but the first

Table 4. Mood-congruent events by area: boredom game level

	total (ext.)	correct	rate
area1	176	165	93.7%
area2	211	207	98.1%
area3	126	116	92.1%
Tot.	513	488	95.1%

Table 6. Mood-congruent events by type: boredom game level

	total (ext.)	correct	rate
dialogue	224	214	95.5%
single fight	136	127	93.4%
chest opened	153	147	96.7%
Tot.	513	488	95.1%

Table 5. Mood-congruent events by area: flow game level

	total	correct	rate
area1	130	124	95.4%
area2	163	159	97.5%
area3	157	150	95.5%
area4	60	59	98.3%
area5	3	2	66.7%
Tot.	513	494	96.3%

Table 7. Mood-congruent events by type: flow game level

	total	correct	rate
dubbed dialogue	46	46	100%
riddle dialogue	18	17	94.4%
dialogue	159	152	95.6%
single fight	153	144	94.1%
fight vs. a group	37	36	97.3%
chest opened	85	84	98.8%
skills upgrade	11	11	100%
stealing action	4	4	100%
Tot.	513	494	96.3%

two have a reduced extent; also here *chest opened* (98.8%) is the best of the activities followed by *group fight* (97%); all classification ratings are over 94% .

6 Conclusions and Future work

The positive results prove that well-designed formal guidelines can manipulate player’s emotions and feelings and that, by using the proposed methodology, good affective experimentations are possible, even with cheap and easy-to-buy headsets; moreover it appears that physiological data as EEG (although with a pre-classification of frequencies) can be effectively treated by the reliable machine learning methods exploited here.

The next step will expand this study to the gamification of real-world serious tasks (medical, teaching), even managing the raw EEG data; analyzing differences between male and female subjects can also help market product-placement.

References

1. Balducci, F., Grana, C., Cucchiara, R.: Classification of affective data to evaluate the level design in a role-playing videogame. In: Games and Virtual Worlds for Serious Applications (VS-Games), 2015 7th International Conf. on. pp. 1–8. Skovde
2. Balducci, F., Grana, C., Cucchiara, R.: Affective level design for a role-playing videogame evaluated by a brain–computer interface and machine learning methods. *The Visual Computer* 33(4), 413–427 (2017), <http://dx.doi.org/10.1007/s00371-016-1320-2>
3. Chanel, G., Rebetez, C., Bétrancourt, M., Pun, T.: Boredom, engagement and anxiety as indicators for adaptation to difficulty in games. In: Proceedings of the 12th International Conference on Entertainment and Media in the Ubiquitous Era. pp. 13–17. MindTrek '08, ACM, New York, NY, USA (2008)
4. Craigh, L.: Game taxonomies: A high level framework for game analysis and design. Gamasutra www.gamasutra.com/features/20031003/lindley_01.shtml
5. Csikszentmihalyi, M.: *Beyond boredom and anxiety*. Jossey-Bass (2000)
6. Frasca, G.: Simulation versus narrative: Introduction to ludology. In: *The Video Game Theory Reader*, pp. 221–236. Routledge, New York (2003)
7. Gee, J.P.: Learning and games. *The ecology of games: Connecting youth, games, and learning* 3, 21–40 (2008)
8. Grimes, D., Tan, D.S., Hudson, S.E., Shenoy, P., Rao, R.P.: Feasibility and pragmatics of classifying working memory load with an electroencephalograph. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 835–844. ACM (2008)
9. Lankoski, P.: Models for story consistency and interestingness in single-player rpgs. In: Proceedings of International Conference on Making Sense of Converging Media. pp. 246:246–246:253. AcademicMindTrek '13, ACM, New York, NY, USA (2013)
10. Lee, J.C., Tan, D.S.: Using a low-cost electroencephalograph for task classification in hci research. In: Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology. pp. 81–90. UIST '06, New York, NY, USA
11. Mandryk, R.L., Inkpen, K.M.: Physiological indicators for the evaluation of co-located collaborative play. In: Proceedings of the 2004 ACM conference on Computer supported cooperative work. pp. 102–111. ACM (2004)
12. Nacke, L., Stellmach, S., Lindley, C.: Electroencephalographic assessment of player experience: A pilot study in affective ludology. *Simulation and Gaming* 42(5), 632–655 (2011)
13. Obbink, M., Gürkök, H., Bos, D.P.O., Hakvoort, G., Poel, M., Nijholt, A.: Social interaction in a cooperative brain-computer interface game. In: *Intelligent Technologies for Interactive Entertainment*, pp. 183–192. Springer (2012)
14. Russell, J.A.: A circumplex model of affect. *Journal of personality and social psychology* 39(6), 1161 (1980)
15. Smith, M.E., Gevins, A., Brown, H., Karnik, A., Du, R.: Monitoring task loading with multivariate eeg measures during complex forms of human-computer interaction. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 43(3), 366–380 (2001)
16. Stieger, I.: Neverwinter nights extender v.4. <http://www.nwnx.org/>, <http://www.nwnx.org/>
17. Vachiratamporn, V., Moriyama, K., Fukui, K., Numao, M.: An implementation of affective adaptation in survival horror games. In: 2014 IEEE Conference on Computational Intelligence and Games, CIG 2014, Dortmund, Germany, August 26–29, 2014. pp. 1–8 (2014)