Maastricht University

e3Learning: A Dataset for Affect-Driven Adaptation of Computer-Based Learning

Citation for published version (APA):

Athanasiadis, C., Amestoy, M., Hortal, E., & Asteriadis, S. (2020). e3Learning: A Dataset for Affect-Driven Adaptation of Computer-Based Learning. *Ieee Multimedia*, *27*(1), 49-60. https://doi.org/10.1109/mmul.2019.2945716

Document status and date:
Published: 01/01/2020

DOI:
10.1109/mmul.2019.2945716

Document Version:
Publisher's PDF, also known as Version of record

Document license:
Taverne

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy
If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Download date: 29 Mar. 2024

e^{3}learning: a Dataset for Affect-driven Adaptation of Computer-Based Learning

Christos Athanasiadis, Matteo Amestoy, Enrique Hortal and Stylianos Asteriadis

Abstract—The use of recommendation systems has been extensively applied to different fields from the suggestion of multimedia content to other areas such as education. Thereof, the integration of affect-related information becomes a key factor to enhance the user experience and, when it comes to learning, it can be translated into the maximization of knowledge acquisition. The existence of affect-enhanced datasets is imperative when training a recommender learning system. This paper introduces a new dataset gathered under realistic conditions which includes affect-related information of 33 learners from different knowledge background interacting with an e-learning platform. Matrix Factorization techniques were evaluated in order to prove the wealthiness of the dataset and to demonstrate that it could potentially replace the adaptation mechanism of the introduced platform and maximizing the learner's knowledge acquisition.

Index Terms—Intelligent learning environments, Computer-based Education, Affective state, Personalized e-learning, Recommendation system.

1 Introduction

Recommendation systems have become a powerful tool in the present-day information society. They have gained increasing popularity in recent years particularly for the advising of multimedia content [1] such as films, music or books or the suggestion of products and services (e-commerce) [2].

On the other hand, recent advances in the area of Affective Computing have enabled the implementation of robust affect analysis tools in several research domains, such as human-computer interaction, computer-based education and activity recognition. Affective analysis on learning platforms is applied in order to analyze emotions emerging from affect information extracted during the interaction of the users with the learning platforms. As customers' preferences are used in classical recommendation systems, users' affective states can be implicitly correlated with their demands and needs and, therefore, the user experience could be enhanced and optimized through personalization and recommendation approaches using emotions as a basis.

In terms of interactive tasks such as e-learning activities, specific emotions experienced by the users may be of particular interest [3] [4]. Learners experience different affective states during the learning process mainly depending on their specific levels of knowledge, and/or competencies and the difficulty of the task at hand. Gathering data containing information regarding this process and the appropriate affective state model could be used for yielding powerful algorithms for affect analysis in real operational learning environments.

• *C. Athanasiadis, M. Amestoy E. Hortal and S. Asteriadis are with the Robots, Agents and Interaction group, Department of Data Science and Knowledge Engineering, Faculty of Science and Engineering, Maastricht University. Address: Bouillonstraat 8-10, 6211 LH Maastricht, The Netherlands.*
E-mail: {c.athanasiadis, matteo.amestoy, enrique.hortal, stelios.asteriadis}@maastrichtuniversity.nl.

For researchers to robustly address the affect analysis task, it is vital to take advantage of datasets to train novel algorithms which will lead to efficient emotion retrieval systems. Much of the progress in the field relies on the availability of public datasets extracted under pragmatic conditions which can be easily handled by a wide spectrum of different algorithms and approaches. For instance, the datasets developed and introduced in [5] [6] helped and pushed the state of the art results in the domain of audio-visual emotion recognition.

In this work, a new learning-related dataset consisting of 33 individuals gathered under realistic conditions was developed. This dataset contains affect-related information following suitable emotional model for some of the most relevant emotions experienced while learning activities are conducted, namely frustration, boredom and flow. The feasibility of this dataset for the development of affect-driven recommendations has been also proved by applying three state-of-the-art matrix factorization techniques. This study is an expansion of a previous one presented in [7] where the learning platform used to capture the dataset that was introduced and described. In juxtaposition with the previous study, the dataset has been enriched with subjects from a different knowledge background (psychology students) and, in addition, more participants have been included to the initial group (from the engineering background). Therefore, the presented database can be considered to be more heterogeneous and complete, also supported by the statistical analysis described in Section 5.1. Moreover, additional experiments were performed with the expanded database using different optimization methods for Matrix Factorization (MF) and also by making use of not only the affect information but the performance of the learners as well (described in Section 6). Extensive investigation using different sparsity levels of available data were used during the MF experiments to showcase the efficiency of the dataset. Additionally, a newly proposed method which

merges the affect states of the learners was introduced. Finally, a new target of the current analysis was to test a system that could potentially propose to learners not only suitable levels of difficulty but also subjects by including all the relevant information in a joint matrix. This system could be integrated in the future to model learners' behaviour during their interaction with e-learning platforms.

2 Related Work

A recommendation system (RS) is a category of information filtering algorithms which aims at automatically suggesting items (such as multimedia content or products) to the users that might be of their interest. Probably, the most popular work is the one conducted by Koren et al. [1] which implements a Singular Value Decomposition-based Matrix Factorization system to create an RS using the Netflix database. Similarly, authors in [8] proposed a probabilistic matrix factorization (PMF) for decomposing the rating matrix of users-items using the same database. But this methodology has been successfully applied not only on the recommendation of multimedia content but also on computer-based education. In that case, the main scope is to model learner profiles [9] to recommend suitable learning content. The main drawback of the previous and similar systems is that they do not take into consideration an important source of information, the user's affective state.

This is also a common handicap for available educational databases and adaptive e-learning mechanisms. Data from the KDD Cup 2010, an educational database [10] which contains questions from algebra topics in several steps and difficulty and parameters related to the learner performance (answer, time spent, etc.) or the ones presented in the review performed in [11] are some examples. Similarly, several works make use of learners profile and performance information in order to personalize the learning content. The main idea behind these works is the modeling of the learners based on their personal skills [9] or their performance [12]. A slightly different approach was followed by [13]. In that case, facial analysis is used to, instead of inferring the emotional state, predict the performance while reading the instruction of the learning activity to be performed.

However, affective states are considered to play an important role during the learning process, since they influence the efficiency of the learners' knowledge acquisition and extensive research has been conducted in that field as highlighted in other works [3] [4]. For example, Bahreini et al. [14] introduced a system called "Framework for Improving Learning Through Webcams and Microphones" (FILTWAM), whose target is real-time emotion recognition of the learners during the interaction with the learning environment. It is capable of recognizing seven basic emotions: sadness, anger, disgust, fear, happiness, surprise and neutral emotion.

Nonetheless, these emotions are hardly associated with the emotions experienced during the learning process. In that context, different models were proposed such as the one presented by Linnenbrink and Pintrichs [15]. In this bidirectional model, the emotion (or mood) influences the perception of the challenge presented which affects the achievement of the goals and, on a reciprocal basis, it has an impact on the mood. This relationship has also been described in more detail in Csikszentmihalyi's Theory of Flow [16], where learner skill and their perception of the task challenge can let them exhibit a variety of affective states. However, it has been argued that not all emotions are relevant to the learning process in identifying optimal learning experiences. D'Mello and Picard conducted a study [17] on the relevance of some affect states to learning and found frustration (or anxiety), boredom and flow to be the most relevant ones to knowledge and other competence. Based on this simplified version of the Theory of Flow, the learning experience can be represented by two dimensions, the skill of the learners and the challenge presented to them. Attaining a balance between these two parameters, a positive effect on the interaction process can emerge which will potentially enhance the knowledge acquisition process.

For that reason, in this paper, we decided to introduce an affective related dataset which captured within the learning context. Our aim was to test RS algorithm and questioning whether they are capable of potentially being used for controlling the adaptation mechanism of the aforesaid learning related platform [7].

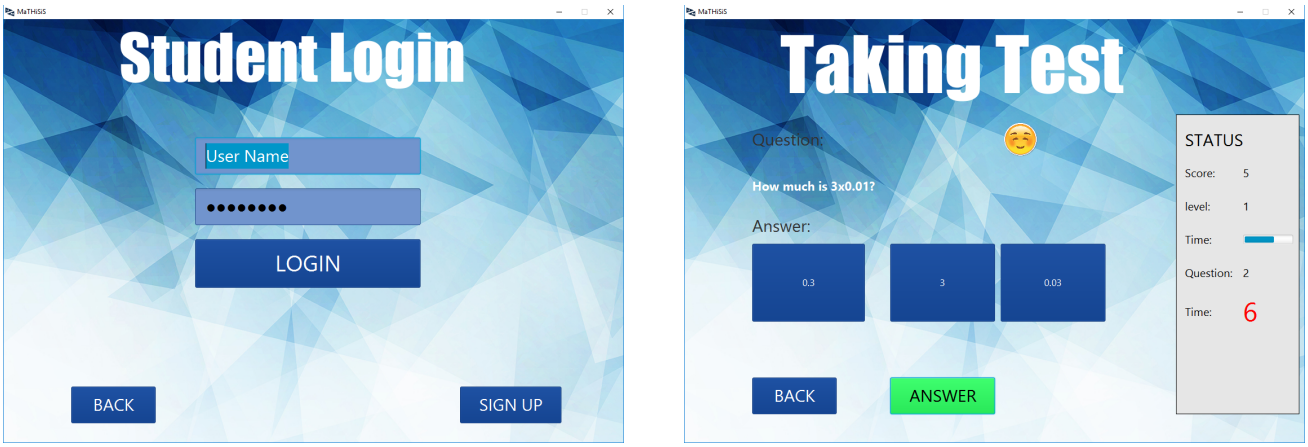
3 The Testbed "e^{3}Learning Platform"

For the research purposes of this study, a new learning platform called *emotion-enabled electronic learning platform* (e^{3}learning platform) was developed (see Figure 1) and is publicly available on Github[1]. This platform consists of two major functionalities, the tutor and the learner scenes. Primarily, the teacher scene was created with a view to facilitating the expansion of the created dataset for future research development. Meanwhile, the main functionality of the platform is the learner scene. This functionality is designed with the aim of gathering the data evaluated in this work and, in the future, to be used as an affect-driven adaptive computer-based education platform.

In the developed version of the learner scene, the learners can select between four different subjects (namely, mathematics, sports, geography and history) and then, up to seven questions per session are consecutively portrayed to them. The presented questions could belong to three different modalities, *choose an answer*, *true or false* and *open answer* and nine levels of difficulty and they are randomly selected from an extensive database provided with the platform. During the learning session, the question, the possible answers (in the case of the two former modalities) and an information box which contains details about the learning session is rendered to the learner. In the case of open answers, learners enter their answers into a rendered box and it is graded as correct or incorrect. Furthermore, different sounds and emoticons are presented in order to enhance the interactivity and provide real-time feedback about the performance of the activity and boost learner's cognitive reactions.

Finally, the result panel interface which informs the learner about the grade obtained during the session and the learner's total grade for the specific subject is displayed.

1. https://github.com/kristosh/e-learning-platform



(a) Learning platform instance 1.

(b) Learning platform instance 2.

Fig. 1: Screenshots from the learning platform.

Further information about the e^{3}learning platform can be found in [7] and in the documentation of the platform[2].

It is important to note here that the adaptation mechanism of the platform was dependent on the score achieved in the previous session. In more details that mechanism is described in 4.

4 Data Acquisition Protocol

After signing a consent form which contains all the information needed to understand the reason behind the experiment and the data acquisition process, the first step implied the students' self-evaluation of their level of competences in the academic subjects included in the study, namely mathematics, sports, geography and history.

Hereafter, the data acquisition protocol (which correspond to platform's adaptation mechanism) was defined in a manner that dataset recorded is rich enough and contains diverse but balance information per user. However, in order to facilitate the experiments' performance and avoid users feeling tired (thus, biasing acquired data), it was decided to establish a restriction per each of the four academic subjects. For this purpose, the nine levels of difficulty available in the platform were grouped into four categories, *Easy* (levels 1 to 3), *Medium* (4-5), *Hard* (6-7) and *Very Hard* (8-9). Then, each learner performed tests belonging to four out of the nine difficulty levels available but covering all the categories. To that end, the approach implemented was dependent on the grade achieved in the previous session. In the beginning, the level from the *Easy* category was established randomly (between levels 1 and 3). Subsequently, the student starts a new session in the next category and the level selected depends on whether or not at least half of the answers were correct in the previous category questionnaire (selecting the highest or lowest level inside the category, respectively).

The whole procedure led to sparse matrices of learners and difficulty levels which are suitable for the study described in Section 7. This protocol allows to record data belonging to all difficulty levels (divided into categories) without exceeding the time constraint, around 30 minutes per participant. Finally, after each learning session conducted, the users were asked to assess, on a Likert scale (from zero to five), the degree of engagement, frustration and boredom that they experienced. This information is used as affect-related data for the analysis performed.

2. https://github.com/kristosh/e-learning-platform/blob/master/README.md

5 e^{3}Learning Dataset

The most important contribution of this study is the introduced publicly available dataset[3] which consists of data gathered from students from Maastricht University, the Netherlands, voluntarily interacting with the e^{3}learning platform described in Section 3. The learners participating were bachelor and master students, from two different knowledge profiles, namely software engineering and psychology. The analysis presented in this paper included 33 users and the duration for the whole experiment was 26$\pm$5 minutes per participant. The dataset architecture was inspired by the works presented in [18] and [19]. Mostly, the information included can be summarized as follows:

- User demographics: personal information such as user name, age and gender, knowledge background (high school, bachelor, master or Ph.D.), the field of expertise (engineering or psychology) and the level of competences in the relevant academic subjects.
- User interactions: information regarding the interaction of the learners with the platform during the learning sessions. This information contains the data related to the questions prompted to each learner (including when they were asked, their subject, level of difficulty, time spent solving the question, or whether or not the correct answer was given), how many times the user has previously interacted with a specific subject, and the emotion-related self-annotation provided at the end of each learning session.

5.1 Dataset demographics & statistics

In this subsection, the demographics and a statistical analyses over the population derived from the dataset are pre-

3. https://project.dke.maastrichtuniversity.nl/kristosh/dataset.rar

sented. The population is distributed as follows: students' age (22.5 ± 3.1), Male/Female (17/16), Bachelor/Master students (22/11) and Engineer/Psychology students (23/10). The features considered in these analyses were grade, time and affective state. In order to investigate the importance of those measurements, independent t-tests were applied to all these populations. The Welch's t-test for unequal variance was chosen to determine whether two independent samples were selected from populations having the same distribution. In order to apply this test, some assumptions needed to be full-filled. Firstly, One-sample Kolmogorov-Smirnov tests were performed with the purpose of testing whether the populations are following or not normal distributions. The result, in all cases, were negative, and consequently, the assumption for normal distribution was met. Furthermore, a second assumption about the independence of the tested populations was certified from the way that the data were captured and analysed (an individual in one group cannot also be a member of the other group and vice versa).

In this paper, we decided to represent only the pairs with the statistical threshold $p < 0.05$. That threshold and the claim of significance have raised the last decades a great debate among the research community about the validity of the statement. However, we decided to proceed with this, since we merely wanted to investigate whether our dataset contains interesting patterns. Some of the pairs of the population where it was found some difference between their distributions (which validated by the P-value of the Welch's tests) along with the median (to showcase the skewness of the distributions), averages values and standard deviation for both populations are rendered in Table 1.

Table 2 includes the mean and standard deviation of the learners' annotation during the learning sessions (namely, boredom, engagement and frustration) on a scale between 0-5 for the two different populations (mean and standard deviation for grade, time and learners' affect self-assessments).

From Table 2, an emerging observation that has been extracted was that the students with engineering background performed and concentrated better in the case of mathematics rather than the students with a background in psychology. The difference between the grade on different students background was validated through Welch's test where the null hypothesis for distribution equality was rejected with $p = 0.05$ while the average grades for engineering/psychology students were 0.62/0.53 respectively. Furthermore, another interest observation that was extracted during Welch's tests was the difference between the self-assessment of male and females for boredom on sports. Female students tended to annotate their affect state with a higher value regarding boredom than male students. On the whole, the analysis presented in the current section revealed interesting observations about the diversity of the students and validates the wealthiness of the dataset.

6 Affect-driven recommendation system

In order to validate the applicability of our approach in the domain of affect-driven computer-based education and its potential application in other fields such as the suggestion of multimedia content, an RS was implemented using the dataset described in Section 5. The concept was to adopt the users' annotations, consider interactions as implicit responses of the system and create a recommendation system using collaborative filtering and, more specifically, Non-Negative Matrix Factorization (NMF) [20], Probabilistic Matrix Factorization (PMF) [8] and Bayesian Matrix Factorization (BMF) [21]. The goal of the experimental study was to develop a system which can adapt the platform's level of difficulty dynamically, exclusively by making use of learners affects, by proposing new levels of difficulty for unseen users, matching them with other learners. The final goal is model a system with the capability to adapt the learning content in pursuit of maximizing learners engagement and, consequently, optimize the learning outcome [17].

First of all, in order to model user's experienced affective states, we introduce a parameter which merges boredom, frustration and engagement levels into a single value, formalized through the use of an energy function. This fusion is based on the hypothesis (aligned with the Theory of Flow) that when learners are in a state of flow, negative entropies like frustration and boredom are not experienced while the engagement of the learners is maximized [22]. Moreover, to move from boredom to frustration the learner has to pass through engagement. Therefore, it seemed rational to establish a scale where the edges correspond to these detrimental states and the full engagement is presented in the middle point. This space could be defined by the following formula:

$$S = \{B, F, E | F \in [0,1], E = 1 - F, B = 0\} \cap \{B, F, E | B \in [0,1], E = 1 - B, F = 0\} \quad (1)$$

Moreover, a way to assign an energy score to each point of S is needed. Assuming that the learner is between boredom and engagement (that is $F = 0$), we want to be able to assign a score that is 0 if the learner is fully engaged and -1 if the learner is fully bored. The simplest way to do that is to define the energy function as follows:

$$\mathcal{E} = \frac{E - B - 1}{2} \quad (2)$$

In the case that the learner is between frustration and boredom we use:

$$\mathcal{E} = \frac{F - E + 1}{2} \quad (3)$$

since it was established that value 1 represents bored students.

As it was mentioned, the aim of this study is to present a mechanism to re-direct learners to a specific **difficulty level** and **subject** in which the engagement will be maximized, therefore, the energy function will be as close to zero as possible. Targeting so, NMF, BMF and PMF were applied in order to optimize the energy function for each learner. In the experimental phase, an aggregated matrix was constructed using gathered information from the four subjects. This matrix contains the energy function values of each user (rows) which represents the corresponding affective state during the performance of each difficulty level and subject (columns). Therefore, the size of the constructed matrix is 33x36, where 33 is the number of learners and 36 is the number of levels (four academic subjects with nine levels of difficulty). In the same vein, a second 33×36 matrix was

Populations	P - Value	Median 1	Median 2	Mean 1	Mean 2	Std 1	Std 2
Geography-Gender-Frustration	0.0009	1.00	2.00	1.42	2.24	1.39	1.44
Sport-Gender-Boredom	0.01	2.00	2.00	1.70	2.38	1.49	1.37
History-Degree-Frustration	0.03	2.00	2.00	1.83	2.53	1.37	1.51
Sports-Gender-Time	0.04	40.5	46.0	42.20	46.91	15.10	10.89
All subjects-Degree-Grade	0.01	4.00	4.00	4.15	3.67	1.79	1.65
All subjects-Faculty-Grade	0.01	4.00	4.00	4.15	3.67	1.79	1.65
All subjects-Gender-Time	0.007	46.00	53.00	50.72	56.26	24.87	23.49

TABLE 1: Some of the results for the Welch's significant tests and the averages for gender, faculty and degree for boredom, engagement and frustration (in the scale of 0-5), time (in seconds) and grade (normalized between 0-1).

Subject	$\mu_{\mathbf{boredom}}(\sigma_{\mathbf{boredom}})$	$\mu_{\mathbf{engagement}}(\sigma_{\mathbf{engagement}})$	$\mu_{\mathbf{frustration}}(\sigma_{\mathbf{frustration}})$	$\mu_{\mathbf{grade}}(\sigma_{\mathbf{grade}})$	$\mu_{\mathbf{time}}(\sigma_{\mathbf{time}})$
Engineering					
Mathematics	0.92 (1.12)	3.00 (1.44)	2.18 (1.49)	0.62 (0.22)	76.80 (23.07)
Sports	1.59 (1.38)	2.02 (1.57)	1.83 (1.49)	0.47 (0.26)	42.96 (14.31)
Geography	1.12 (1.17)	2.56 (1.46)	1.64 (1.42)	0.58 (0.27)	31.04 (12.94)
History	1.30 (1.37)	2.47 (1.49)	2.18 (1.49)	0.56 (0.24)	54.70 (18.43)
Psychology					
Mathematics	1.02 (0.88)	3.41 (1.29)	2.59 (1.26)	0.53 (0.21)	81.63 (19.53)
Sports	1.53 (1.36)	2.10 (1.32)	2.10 (1.31)	0.46 (0.23)	48.15 (10.16)
Geography	1.26 (1.19)	3.23 (1.43)	2.19 (1.61)	0.57 (0.24)	33.24 (9.78)
History	1.00 (0.97)	2.98 (1.44)	2.59 (1.49)	0.54 (0.25)	63.17 (17.38)

TABLE 2: Mean and standard deviation for learners' cognitive self-assessments (boredom, engagement and frustration in the scale of 0-5), time (in seconds) and grade (correct answers normalized between 0-1) for the engineering and psychology populations.

constructed using the grade obtained by the learners in each learning session as a secondary approach, more similar to traditional multimedia RS based on ratings.

Therefore, MF approaches were applied to the two different aforementioned matrices. The endeavour of those algorithms is, given the matrix $R^{M\times N}$ (which corresponds to M users and their annotation or performance to N different learning materials), to find out two matrices $W^{M\times K}$ and $H^{K\times N}$ with their multiplication $A \approx W \cdot H$ to be a good approximation of the initial matrix R. Matrices W and H are correlated with learners and learning materials (academic subjects and their associated difficulty level). MF approaches are based on the following principle: both the user and the learning materials from matrix R should be represented in the same space. On that basis, MF algorithms map learners and learning materials into a common space *k* or latent factors. A high correlation between learning materials and user latent factors can lead to more efficient material recommendations. The matrix approximation A can be used in order to make recommendations to the learner for unknown learning materials since it reveals information about them.

In order to decompose matrix A into W, H as it mentioned already, three different approaches were tested namely NMF, PMF and BMF.

The re-decomposed matrix $A \approx WH$ (in all cases, NMF, PMF and BMF) can reveal hidden values of energy function (or grade) in unknown learning materials for learners in the dataset, as well as, for new learners interacting with platform by making use of the provided user self-annotation, to estimate the energy function (or grade) values for the unknown materials. High-level wise, this estimation is done by matching the new learner's information with the one from well-known learners. The MF algorithms are applied using the constructed aggregated matrix from the dataset, enriched with the new learner's vector. The recommendation for a new learner, then, can be formalized as follows:

$$a_{ik} = w_{ik} * h_{kr} \tag{4}$$

where the w_{ir} vector contains the latent factors of the new learner and the h_{rk} vector contains the latent factors of the learning materials. Eq. 4 predicts learner *i* energy function value or grade for a specific material *r*. Finally, by using this information, the system could potentially re-direct the learner to that material in which the grade value or the engagement for the affect-based matrix will be optimized.

7 Experimental results and discussion

This section outlines striking observations that emerged during the experimental phase. By making use of the proposed dataset and state-of-the-art recommendation algorithms, the efficiency of the dataset and the affect-driven mechanism was established. Experiments using MF were carried out for the aggregated matrices with four academic subjects (namely, mathematics, sports, geography and history) which were constructed using the affect (potentially transferable to other application fields) and alternatively, grading information of the learners (similarly to the scores provided in multimedia content RS). As mentioned in Section 5, each participant interacted with four out of nine difficulty levels for each academic subject. Therefore, in the aggregated matrices each learner interacted with 16 out of 36 of all possible learning materials. Two aggregated matrices (33×36) were obtained using affect and grade information. During the experimental phase, 33-fold cross-validation (leaving one user out per fold) was implemented. In each loop of this cross-validation, the size of the training set was set to 32 samples (32×36 matrix) while the validation set was contained one sample (1×36 vector).

Since the target of the experiments was to determine whether or not the MF algorithms can be used for efficiently

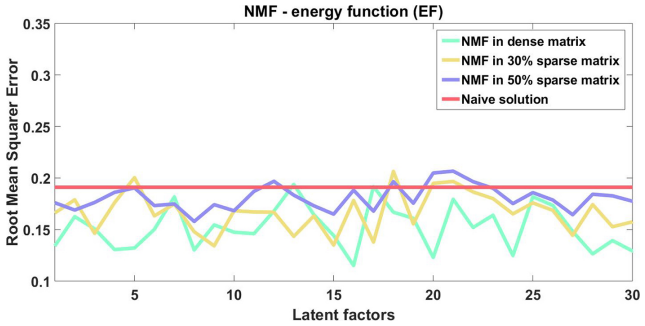
predicting the entry values of the two matrices for unknown learning materials for a specific learner using information from previous interactions, an investigation of the capability of the system to predict those unknown values was performed. For that purpose, one entry from the validation set was randomly removed (from the two matrices). This reduced vector contains information of 15 out of the 16 user's sessions. Subsequently, NMF, BMF and PMF were applied using as decomposition matrix the concatenation of the training and evaluation sets (after the random value removal). The decomposition was applied for all the steps of the cross-validation. Thereby, the scope of the whole procedure was to infer the missing value.

Similarly, to demonstrate the efficiency of the employed MF algorithms over the introduced dataset, further tests with different sparsity levels of both matrices were also performed. In those cases, 30% and 50% of the entry values (per row) were removed and the same evaluation was conducted. These tests were performed to validate the existence of patterns within the dataset even when we limited the amount of data to 30% and 50%.

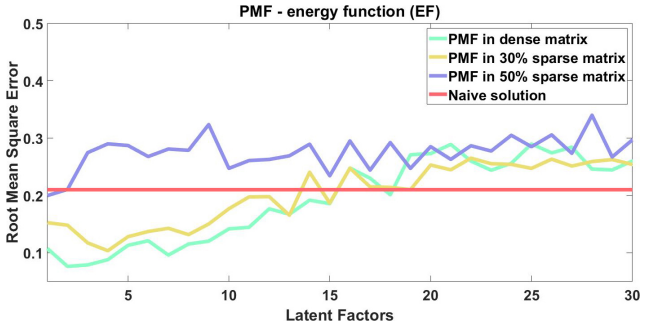
As a prediction indicator of the approximation, the average root mean square error (RMSE) was employed (for all the cross-validation loops), between the missing values from the validation set and the corresponding values contained in the initial evaluation matrix. Fig. 2 and 3 render this root mean square error for all the algorithms and the three different sparsity levels for several values of rank k, after the performed cross-validation. For comparison purposes, a baseline method (marked as a red line in the aforementioned figures) was introduced. This method consists of calculating the root mean square error between the missing value and the average value of each matrix. An exhaustive search was performed to calculate the best parameters of MF algorithms for the validation procedure.

It is interesting to note that, in most cases, for any method, objective function and matrix we can find a latent space dimension that has a better RMSE than the naive solution. This shows that the dataset has a structure that intelligent algorithms can infer. From the comparison of the three methods on the two matrices, we can subtract that the NMF has a nearly constant RMSE over the latent space dimension (leading to similar predictions regardless of the dimension), whereas PMF and BMF reach a minimum in the first dimensions. Whereas, PMF and BMF present usual learning curves: decreasing or constant RMSE for the first dimensions, and then, the model over-fitting makes the RMSE increase.

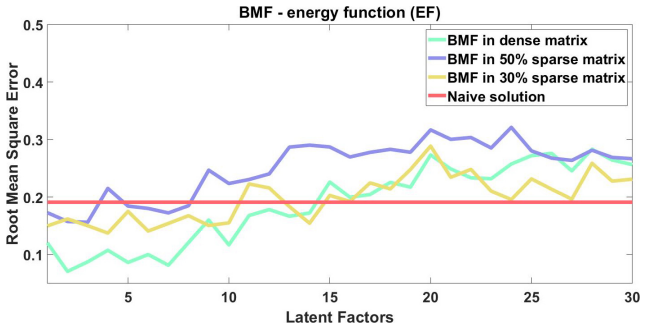
A remarkable emerging observation from the experimental results was the fact that we can successfully implement different MF techniques (NMF, PMF and BMF) to develop robust personalization algorithms for the learning platform introduced in this study. From Fig. 2 and 3, it could be seen that the system can achieve small RMSE and therefore, it could robustly estimate unknown energy values and learners' grade. MF algorithms could be successfully applied as an RS for learning environments similar to the e^{3}learning platform (and therefore to different computer-based education systems) by estimating learners affect or grade and providing personalized learning paths optimizing the selection of learning materials based on this information. Another



(a) Non-negative Matrix Factorization.



(b) Probabilistic Matrix Factorization.



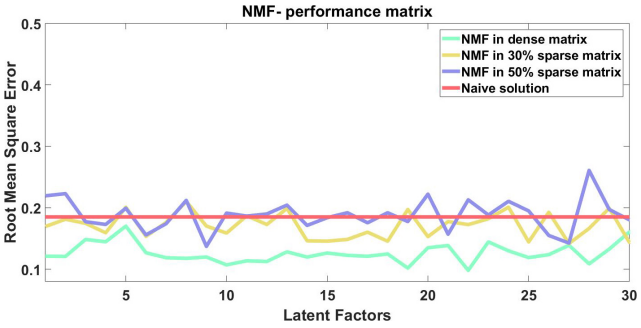
(c) Bayesian Matrix Factorization.

Fig. 2: Root mean square error using different MF algorithms in the matrix containing energy function values for different sparsity levels.

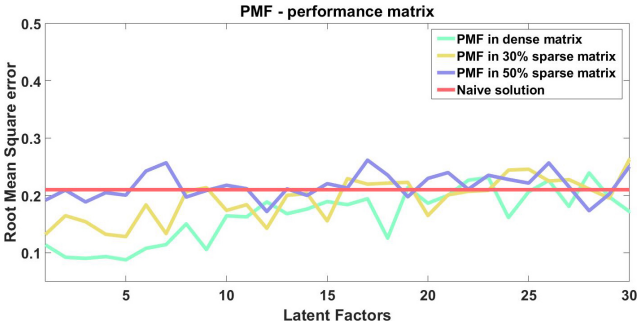
observation emerged concerning the experimental results is that the affective states and learners interactions registered in the dataset were sufficient and rich enough to implement a efficient personalization system using MF, since any missing information (the randomly removed value for the sake of this study) could be robustly modeled and approximated by solely using the knowledge of the learning analytics or annotations of the users. Thereby, we can conclude that our dataset contains a significant plurality of affects and learners' interactions.

7.1 Limitations

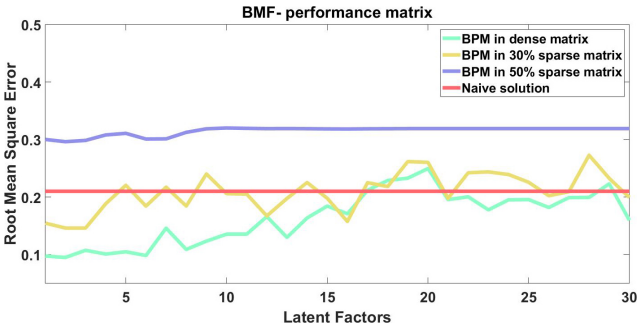
The work presented shows great potential for the dataset and the adaptation mechanism proposed. However, some limitations must be highlighted. Firstly, even though it was proved that the dataset is sufficient and rich enough, the number of users and the data available per user could be limited to develop systems using more demanding algo-



(a) Non-negative Matrix Factorization.



(b) Probabilistic Matrix Factorization.



(c) Bayesian Matrix Factorization.

Fig. 3: Root mean square error using different MF algorithms in the matrices containing energy function values for different sparsity levels.

rithms. However, the publicly available e-learning platform designed to gather this data can be easily used and expanded to enlarge this dataset. Another remarkable downside is the use of the self-perception of affective states in the data labelling process. It leads to some limitations such as human bias, forgetting effects or different interpretations of these affective states. To minimize these effects, we tried to clearly define a set of rules provided both verbally and in writing in the consent form. Finally, the energy function assumes that each of the affective states has a similar impact on individuals' performance. However, it could vary at an individual-level or given the context of the tasks. Further investigations in this area should be performed to define the most appropriate weight for the emotions into consideration.

8 Conclusion

The main contribution of this paper is the presented dataset derived from users' interactions with a learning platform, the so-called e^{3}learning platform, captured under realistic in the wild conditions. There are two important differences between the introduced dataset in comparison with the ones currently available. Firstly, the fact that, up to our knowledge, this is the only dataset that offers the possibility of a linkage between learners levels of engagement, boredom and frustration with their semantic information during the learning sessions. Moreover, the architecture of the dataset and the e-learning platform was designed to offer a user-friendly chance of expanding and enriching the dataset, by adding new learning material to the platform.

Secondly, the dataset offers the possibility of performing recommendations by incorporating features from learners' affective state, their performance and the contextual information (learner's profile and learning material information). The validity of the dataset studied by applying it in an RS framework is another important contribution. From the experimental results, emerging indicators show the capability of our dataset and the collaborative filtering techniques evaluated when used for generating reliable personalization algorithms to be applied in the design of affect-driven adaptable computer-based education platforms. To our knowledge, this is the first time that the combination of collaborative filtering algorithms and affective states as explicit preferences are implemented. Additionally, it is noteworthy that the method aggregates the information from different academic subjects together in a single matrix to find hidden correlations among them and on top of that, different MF methods were applied in the aggregated matrix in an attempt to recommend multimedia content for the learning platform.

It is for all these enhancements that this affect-driven system is intended as a first step for the use of affect-related information in widely used fields. We strongly believe that this affect-driven mechanism can easily replace the adaptation mechanism of the learning platform, as well as, to be applied in other fields such as the recommendation of multimedia content or other products and services. Using this mechanism, these applications will benefit from the affect-related information improving the user experience and providing content not only based on slowly changing users' preferences but also their varying affective state.

References

[1] Y. Koren, R. Bell and C. Volinsky, *Matrix factorization techniques for recommender systems*, Computer, 8, 2009.
[2] J.B. Schafer, J. Konstan, and J. Riedl, *Recommender systems in e-commerce*, in Proceedings of the 1st ACM conference on Electronic commerce, 1999.
[3] R. Pekrun, T. Goetz, W. Titz and R.P. Perry, *Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research*, Educational psychologist, 37(2), 91-105, 2002.
[4] O. Korn and A. Dix, *Educational playgrounds: how context-aware systems enable playful coached learning*, Interactions, 24(1), 54-57, 2016.
[5] A. Dhall, R. Goecke, S. Lucey and T. Gedeon, *Collecting Large, Richly Annotated Facial-Expression Databases from Movies*, IEEE MultiMedia, 19(3), 34-41, 2012.

[6] H. Cao, D.G. Cooper, M.K. Keutmann, R.C. Gur, A. Nenkova, and R. Verma, CREMA-D: Crowd-sourced Emotional Multimodal Actors Dataset, IEEE Transactions of Affective Computing, 5(4), 377390, 2014.
[7] C. Athanasiadis, E. Hortal, D. Koutsoukos, C.Z. Lens and S. Asteriadis, *Personalized, affect and performance-driven Computer-based Learning*, in Proceedings 9th CSEDU, 2017.
[8] A. Mnih, and R.R. Salakhutdinov, *Probabilistic matrix factorization*, in Advances in neural information processing systems, 2008.
[9] J. Bobadilla, F. Serradilla and A. Hernando, *Collaborative Filtering Adapted to Recommender Systems of e-Learning*, Knowledge-Based Systems, 22(4), 2009.
[10] J. Stamper, A. Niculescu-Mizil, S. Ritter, G.J. Gordon and K.R. Koedinger, *Data set from KDD Cup 2010 Educational Data Mining Challenge*. Find it at http://pslcdatashop.web.cmu.edu/KDDCup/downloads.jsp
[11] K. Verbert, N. Manouselis, H. Draschler and E. Duval, *Dataset-Driven Research to support learning and knowledge Analytics*, Educational Technology and Society, 15(3), 2012.
[12] E. Lavoué, B. Monterrat, M. Desmarais and S. George, *Adaptive gamification for learning environments*. IEEE Transactions on Learning Technologies, 2018.
[13] C. Lopez, and C. Tucker, *Towards personalized performance feedback: mining the dynamics of facial keypoint data in engineering lab environments*, in ASEE Mid-Atlantic Section Spring conference, 2018.
[14] K. Bahreini, R. Nadolski and W. Westera, *Towards multimodal emotion recognition in e-learning environments*, Interactive Learning Environment, 24(3), 2016.
[15] E.A. Linnenbrink, and P.R. Pintrich, *Achievement goal theory and affect: An asymmetrical bidirectional model*, Educational psychologist, 37(2), 2002.
[16] M. Csikszentmihalyi,*Flow: The psychology of optimal experience*. London: Harper and Row, 1990.
[17] S. D'Mello, R.W. Picard and A. Graesser, *Toward an affect-sensitive AutoTutor*, IEEE Intelligent Systems, 22(4), 2007.
[18] N. Shaker, S. Asteriadis, G.N. Yannakakis and K. Karpouzis, *Fusing Visual and Behavioral Cues for Modeling User Experience in Games*, IEEE transactions one Cybernetics, 43(6), 2013.
[19] N. Shaker, S. Asteriadis, G.N. Yannakakis and K. Karpouzis, *A Game-based Corpus for Analyzing the Interplay between Game Context and Player Experience*, Affective Computing and Intelligent Interaction: Fourth International Conference, 2011.
[20] D. Lee and H.S. Seung, *Learning the parts of objects by non-negative matrix factorization*, Nature, 401, 1999.
[21] R. Salakhutdinov and A. Mnih, *Bayesian Probabilistic Matrix Factorization using Markov Chain Monte Carlo*, in Proceedings of the 25th International Conference on Machine Learning, 2008.
[22] J. Nakamura and M. Csikszentmihalyi, *The concept of flow*, Emotion focused approaches, 2004.