Towards entertaining and efficient educational games

András Lőrincz
Faculty of Informatics
Eötvös Loránd University
Budapest, Hungary 1117

Gyöngyvér Molnár
Institute of Education
University of Szeged
Szeged, Hungary 6722

László A. Jeni
Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213

Zoltán Tősér
Faculty of Informatics
Eötvös Loránd University
Budapest, Hungary 1117

Attila Rausch
Institute of Education
University of Szeged
Szeged, Hungary 6722

Jeffrey F. Cohn
Department of Psychology
University of Pittsburgh
Pittsburgh, PA 15213

Benő Csapó
MTA-SZTE Research Group on the
Development of Competencies
University of Szeged
Szeged, Hungary 6722

Abstract

We have been developing and combining components of an adaptable and entertaining educational gaming framework. To optimize individual learning trajectories, we need to estimate motivations, emotions, cognitive capabilities and progress of learners. Our framework encompasses real-time facial expression estimation and feedback, a prototype training game set on inductive reasoning, and a large battery of tests. Here, we report results with 180 elementary school students. We studied our framework in different conditions and found that it contributes to optimal cognitive development. For illustrative purposes, we show computer detectable signs of expectations concerning success and predictive signs of comprehension of test requirements.

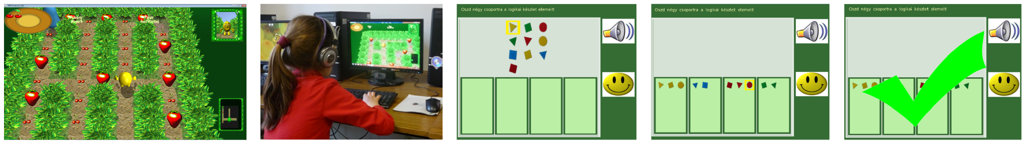
1 Introduction

The cumulative disadvantage of children in rural areas has been a long-standing problem: the environment, teachers, distance to schools, developing concepts in time, such as the mental number line are of serious problems. Talents may also be accompanied by special needs that may be in conflict with the possibilities. Inadequate knowledge, the lack of an experience based and coherent view of our world have far reaching consequences for economic and social wellbeing [6]. We are looking for early educational methods that enable children to learn to their maximum ability, at the most effective and entertaining way, at a much lower cost and early enough.

Interactive games can attract children, entertain and induce learning, and can be adapted with respect to the child during playing. We have been developing such games together with evaluation and recommendation methods in order to keep the child as close as possible to the 'flow' [3].

2 Methods

We trained first grade students attending one of the elementary schools in Szeged, Hungary with an educational game. The structure of the training was based on Klauer's [7] original inductive reasoning training program with the six basic classes, such as generalization, discrimination, cross classification, recognizing relationships, differentiating relationships, and system construction.



(a) (b) (c) (d) (e)

Figure 1: Pac-Man like framegame and minigame. (a): yellow and blue characters: player and ghost, cherries: worth of points, strawberries: positions to start minigames (cf. power pills), character in upper-right corner: feedback reflecting behavior of player (happy: player faces the game, getting angry: player is not facing the game), (b): a player, (c): a task, (d): a proper answer, (e): feedback about the result.

2.1 Minigames and framegame embedding

The computer game contains 120 problems, that is, in computer-based environment 120 minigames made of 20 minigames in each class of inductive reasoning. In the minigames we used pictures, problems and stories that correspond to the interest of children today. The tasks can be solved through the application of appropriate inductive reasoning processes [9]. Minigames were designed so that students perceive the training process as playing instead of learning. Students used headsets and listened to the tasks, so tasks require no reading skills. Students worked individually and each task started with a problem statement; a description of the minigame. Two or three facilitators were present and helped in case of computer problems during the training. Students played about 20 minutes twice a week and they had 12 minigames to solve on each occasions. At the end of every minigame they received audio-visual feedback about whether their answer was right or wrong. In the case of failure the minigame started again. For more information about the training, see [2].

We integrated the minigames into a Pac-Man like arcade game (Fig. 1). To make the game suitable for children we changed the traditional visual form of Pac-Man: walls were bushes, dots were cherries and power-pills were strawberries. In the framegame children could collect cherries and strawberries for points. Picking up a strawberry opens a minigame. After a successful solution the child was able to chase and knock out the ghosts for a while. In case of failure, the ghosts chased them and pushed them back to the start. Experiments were conducted with and without the arcade game, but in both cases avatar based visual feedback was added: the head pose and facial expressions were monitored and an avatar on the screen became increasingly angry and more-and-more red-faced if the child did not look at the computer, but relaxed if the child was looking at the screen.

2.2 Facial Feature Point Localization

For monitoring and facial expression evaluations, we used a dense set of facial landmarks provided by LiveDriver$^{\mathrm{TM}}$[1] software. This tracker registers 64 2D facial feature points to an image. Using a non-rigid structure-from-motion algorithm [12] we constructed the corresponding 3D shape modes of the 2D marker-set.

The 3D shape model is defined by a 3D mesh; vertices of the mesh are called landmark points and the shape is made of their 3D coordinates $\mathbf{x} = [x_1; y_1; z_1; \ldots; x_M; y_M; z_M]$, or, $\mathbf{x} = [\mathbf{x}_1; \ldots; \mathbf{x}_M]$, where $\mathbf{x}_i = [x_i; y_i; z_i]$. We have T samples: $\{\mathbf{x}(t)\}_{t=1}^{T}$. We assume that – apart from scale, rotation, and translation – all samples $\{\mathbf{x}(t)\}_{t=1}^{T}$ can be approximated by means of the linear principal component analysis (PCA). Next, we briefly describe the 3D Point Distribution Model.

2.2.1 Point Distribution Model

The 3D point distribution model (PDM) describes non-rigid shape variations linearly and composes it with a global rigid transformation, placing the shape in the image frame:

$$\mathbf{x}_i(\mathbf{p}) = s\mathbf{P}\mathbf{R}(\bar{\mathbf{x}}_i + \mathbf{\Phi}_i\mathbf{q}) + \mathbf{t}, \tag{1}$$

[1]LiveDriver$^{\mathrm{TM}}$ SDK for Windows, Version [1.0]. Copyright© 2012 Image Metrics, see `http://www.image-metrics.com/livedriver`

$(i = 1, \ldots, M)$, where $\mathbf{x}_i(\mathbf{p})$ denotes the 3D location of the i^{th} landmark and $\mathbf{p} = \{s, \alpha, \beta, \gamma, \mathbf{q}, \mathbf{t}\}$ denotes the parameters of the model, which consist of a global scaling s, angles of rotation in three dimensions ($\mathbf{R} = \mathbf{R}_1(\alpha)\mathbf{R}_2(\beta)\mathbf{R}_3(\gamma)$), a translation $\mathbf{t}$ and non-rigid transformation $\mathbf{q}$. Here $\bar{\mathbf{x}}_i$ denotes the mean location of the i^{th} landmark (i.e. $\bar{\mathbf{x}}_i = [\bar{x}_i; \bar{y}_i; \bar{z}_i]$ and $\bar{\mathbf{x}} = [\bar{\mathbf{x}}_1; \ldots; \bar{\mathbf{x}}_M]$) and $\mathbf{P}$ denotes the projection matrix to 2D. We assume that the prior of the parameters follow a normal distribution with mean $\mathbf{0}$ and variance $\mathbf{\Lambda}$ at a parameter vector $\mathbf{q}$: $p(\mathbf{p}) \propto N(\mathbf{q}; \mathbf{0}, \mathbf{\Lambda})$ and we used PCA to determine the d pieces of $3M$ dimensional basis vectors ($\mathbf{\Phi} = (\mathbf{\Phi}_1^T, \ldots, \mathbf{\Phi}_M^T)^T \in \mathbb{R}^{3M \times d}$). Vector $\mathbf{q}$ represents the 3D distortion of the face in the $3M \times d$ dimensional subspace and it can be used for emotion classification, for example.

We used the BU-4DFE dataset [13] to construct the 3D PDM and to register face images, 3D structure from motion was estimated using the method of Xiao et al. [12].

2.3 Global Alignment time-series kernel

Kernel based classifiers, like any other classification scheme, should be robust against invariances and distortions, including dynamic time warping, which is traditionally solved by dynamic programming. Recently, efficient kernel methods have appeared [4]. We used the Global Alignment (GA) kernel in our studies. GA kernel assumes that the minimum value of alignments may be sensitive to peculiarities of the time series and replaces this sensitive quantity with the sum of all alignments weighted exponentially:

$$k_{GA}(x, y) \stackrel{def}{=} \sum_{\pi \in A(n,m)} e^{-D_{x,y}(\pi)}. \tag{2}$$

Equation (2) can be rewritten by breaking up the alignment distances according to the local divergences: $k_{GA}(x, y) \stackrel{def}{=} \sum_{\pi \in A(n,m)} \prod_{i=i}^{|\pi|} e^{-\phi(x_{\pi_1(i)}, y_{\pi_2(i)})}$. According to the argumentation, since k_{GA} runs over the whole spectrum of the costs it gives rise to a smoother measure than the minimum of these costs. The computational effort is $\mathcal{O}(nm)$ and, importantly, the induced Gram matrix do not tend to be diagonally dominated as long as the temporal sequences have similar lengths [4]. We used kernel $e^{-\phi_\sigma}$ with $\phi_\sigma \stackrel{def}{=} \frac{1}{2\sigma^2}||x - y||^2 + \log\left(2 - \exp\left[-||x - y||^2/(2\sigma^2)\right]\right)$ [4, 8].

We utilized the thresholded kernel similarity measure in one of our illustrative example. In the other one, we used Support Vector Machine classifier [5].

3 Experimental Results

We have monitored three training courses during the spring of 2011 (1 group), and the spring and fall of 2012 (two-two groups). In each of the five groups we had close to 30 children who finished the course. A few drop-outs occurred due to illnesses for longer periods. The present examples are chosen from the dataset that includes time series extracted from videos of 57 subjects playing games during the fall of 2012. In one of our analyses, we have extracted 2 seconds long segments from the videos, and used the corresponding PCA parameters as temporal variables. In the other cases, we used raw 3D data of landmarks around the mouth. Our analysis concerned the success or failure of the trials.

The effectiveness of the training was measured with an inductive reasoning test, which consists of 37 non-verbal items. The Cronbach α reliability coefficient of this test is 0.87. For pre- and post-testing we used our online electronic diagnostic assessment platform[2]. Testing sessions and developmental training took place at the same computer rooms of the school. Testing sessions lasted approximately 45 minutes and they were supervised by teachers. Students achieved 24,0% on the pre-test and 42,2% on the post-test (t=18,2; p<0,001), their inductive reasoning skills developed by over 18%. They significantly outperformed the control group by more than one standard deviation (Fig. 2). Their performance lasted out, six month later they achieved 44,4% on the follow-up test. There were no significant differences between the two groups (frame game + minigames, minigames). After the training the pupils were asked about the game, 94,8% liked it very much and would like to play this game another time, if they had a chance.

The standard deviation of performance increased from the pre- to the post-test in all of the scenarios with one exception: when the minigames were embedded into an extra gaming environment,

[2] eDia, see `http://edia.edu.u-szeged.hu/`

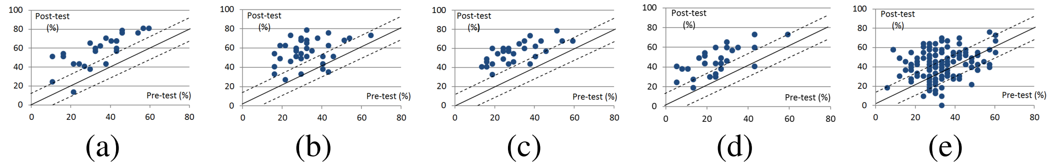


Figure 2: Pre-test and post-test results for children. (a): Playing mostly alone, (b): situation is more social, children can communicate occasionally, (c): playing mostly alone and frame game is added, (d): as (c), but with avatar feedback, (e): control groups in one figure.

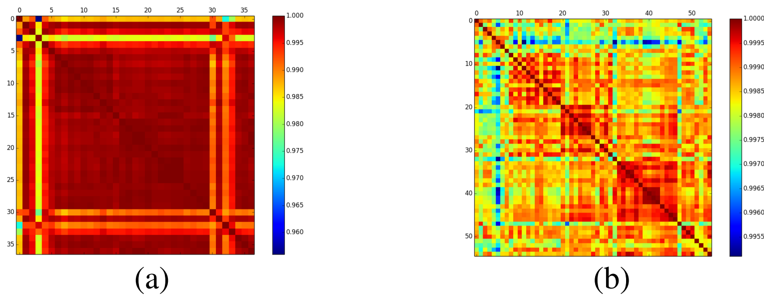


Figure 3: Gram matrices for dynamic expressions. Kernel similarity measure is depicted by colors. Value 1 means perfect similarity. (a): Failures induce very little facial expressions, unless a positive result is expected. Indices 0, 2, 3, 4, 5, 30, 32, and 33 correspond to the special expression. (b): Examples with positive (first eight indices) and negative feedback with the exception of the above indices.

students' achievement were more similar after the training sessions. Generally, in all of the piloted environments the training program increased students' learning outcomes by more than one standard deviation. A detailed study of the different conditions is in preparatory phase.

We show two demonstrative cases. In the first one, the facial expression response tells if a good answer was expected, but it was bad. In the second one, the facial expression predicts much earlier if the child does not know the answer. These examples were selected since they represent cases when a good teacher or playing partner would interact. Both examples describe rare cases.

3.1 Case Example 1

In this case, the pupil thinks that the solution should be good, but it is not. The expression is maintained during the short feedback intervals and similar expressions occur in other contexts, too. This is a special and mixed expression that involves only the mouth, has components of anger and surprise. Figure 3(a) shows the Gram matrix with 36 instances out of which only 8 has this expression in a video which lasts for an hour. The intervals were about 2s long only. The strength of the expression is sufficient for shape-based analysis. The Gram matrix was created with GA kernel and for the shape of the mouth, i.e., for 26 3D marker points. Frame rate varied due to simultaneous facial feature point tracking and were around 15 fps. Some of the examples stand out (they are visible as vertical and horizontal lines), whereas some others are barely visible. It is easy to mix this expression with the case when the mouth opens, or if the child is murmuring. Figure 3(b) shows the Gram matrix for the examples where we took out the 8 samples and included the positive feedback cases. They are mostly neutral, but on some of them – that we placed to first 8 indices – a slight smile becomes visible. The difference can be seen by inspecting the relatively bluish part of the first 8 columns outside of the 8×8 upper right square. Red squares represent examples that belong to the same video session.

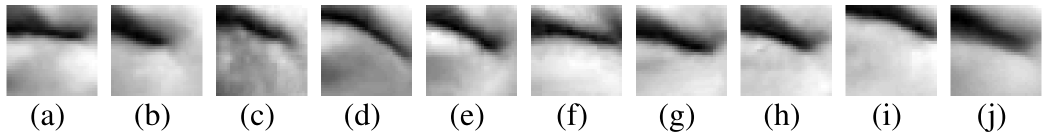


Figure 4: Contempt like predictive expressions. Patches taken around the lower-left-middle marker point of the mouth. (a): neutral, (b)-(i): predictive expressions, (j): average of the (b)-(i) patches

3.2 Case Example 2

This example is of another case about another child. In this case, there were times that the child did not understand the problem. The facial sign is a slightly squeezed mouth that sometimes curves down too. The expression is somewhat similar to contempt and anger. It is weak and sometimes very short. We cut out patches around one of the markers, see Fig. 4. The first inset shows the same patch for a neutral face. The variation between patches is relatively large, information is both in the shape and in the texture. For both the examples, special methods are needed to find them. Similarity measures built onto the positive samples might be a viable route [1].

As we have already noted, standard deviation of performance increased from the pre- to the post-test in all of the scenarios with one exception: when the minigames were embedded into an extra gaming environment, students achievement were more similar after the training sessions. Visual inspection of Fig. 2(b) points to the possibility that a more social environment improves the pre- to post-test progress of children with lower pre-test scores. Further experiments are in preparation.

3.3 Discussion of Case Examples

We showed two illustrative examples for learning scenarios. An experienced teacher would have interacted in both cases to optimize the learning process. In our case, the computer could also interact and managed to interpret correctly at an estimated 50% and 20% in the first and the second examples, respectively. This is already promising, but the rate could be much higher if longer histories, including the number of preceding failures were known. Furthermore, the facial expression following feedback is mostly neutral in the first example, no matter if the result is failure or success. For a longer series of failures, the interaction is necessary to maintain motivation and interest.

3.3.1 Two stage recommendations

Training curriculum can be optimized to the goals and to the subject. It depends on the learning history, motivations and abilities. Learning tasks should be within the *zone of proximal development* [11] and should cover at least the relevant topics for reaching the goals. Individual learning trajectories matched to the characteristics of the actual learning provide the examples and can serve recommendations. This is called *collaborative filtering* and methods are subject to extensive research and are steadily improving (see, e.g., [10] and the references therein).

Learning improves when one can focus efficiently and effortlessly [3]. This is a highly demanding desire, but appropriate steps should be taken in case of frustration when the task is not in the proximal zone (our first example), or if the material is in that zone, but the task is not understood (our second example).

Further problems arise when the child is in the flow since – according to experiences and our recordings, too – facial expression can be blunt and one can easily confuse it with other situations like ‘being inattentive’. In turn, novel and verifiable methods are in need for educational games. This is a big challenge, but might be feasible since a well designed educational game may restrict the world to that of the game.

4 Conclusions

Efficient, personalized, and entertaining computer based educational games are of high importance for the society. we have provided examples for such games. Training results are robust against environmental conditions and pupils liked it. Nonetheless, high level of entertainment remains a highly demanding challenge since individual emotional signs should be recognized and then exploited in the actual context. Such signs can be very short, are diverse, subtle, rare, and may depend on culture. Games that provoke emotional facial expressions could alleviate the situation. Since a more social environment did not spoil progress we suggest that multi-player cooperative-competitive games might be appropriate to overcome the 'mind-reading' problem.

Acknowledgments

This research was funded in part by the European Union (Project ID number: TÁMOP-3.1.9-11/1-2012-0001) and it is also part of the EIT KIC 12-1-2012-0001 project, which is supported by the Hungarian Government, managed by the National Development Agency, financed by the Research and Technology Innovation Fund and was performed in cooperation with the EIT ICT Labs Budapest Associate Partner Group.

References

[1] A. Bar-Hillel, T. Hertz, N. Shental, and D. Weinshall. Learning a mahalanobis metric from equivalence constraints. *Journal of Machine Learning Research*, 6(6):937–965, 2005.

[2] B. Csapó, A. Lőrincz, and G. Molnár. Innovative Assessment Technologies in Educational Games Designed for Young Students *Assessment in game-based learning. Foundations, innovations, and perspectives*, pages 235–254. Springer, New York, NY, 2012.

[3] M. Csikszentmihalyi. Finding flow: The psychology of engagement with everyday life. *Basic Books, 1997*.

[4] M. Cuturi. Fast global alignment kernels. In *Proceedings of the International Conference on Machine Learning*, 2011.

[5] L. A. Jeni, A. Lőrincz, T. Nagy, Z. Palotai, J. Sebők, Z. Szabó, and D. Takács. 3d shape estimation in video sequences provides high precision evaluation of facial expressions. *Image and Vision Computing*, 30(10):785–795, 2012.

[6] J. Johnson, S. B. Shum, A. Willis, S. Bishop, T. Zamenopoulos, S. Swithenby, R. MacKay, Y. Merali, A. Lőrincz, C. Costea, et al. The FuturICT education accelerator. *The European Physical Journal Special Topics*, 214(1):215–243, 2012.

[7] K. J. Klauer. Denktraining für Kinder. *Hogrefe, Göttingen*, 1989.

[8] A. Lőrincz, L. A. Jeni, Z. Szabó, J. F. Cohn, and T. Kanade. Emotional Expression Classification Using Time-Series Kernels. In *Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 889–895. 23-28 June 2013

[9] G. Molnár. Playful fostering of 6- to 8-year-old students' inductive reasoning. *Thinking Skills and Creativity*, 6(2):91–99, 2011.

[10] Z. Szabó, B. Póczos, and A. Lőrincz. Collaborative filtering via group-structured dictionary learning. In *Latent Variable Analysis and Signal Separation*, pages 247–254. Springer, 2012.

[11] F. R. Volkmar. Zone of proximal development (zpd). In *Encyclopedia of Autism Spectrum Disorders*, pages 3421–3421. Springer, 2013.

[12] J. Xiao, S. Baker, I. Matthews, T. Kanade. Real-time combined 2D+3D active appearance models. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Washington, D.C., USA, 535–542, 2004.

[13] L. Yin, X Chen, Y. Sun, T. Worm, M. Reale. A high-resolution 3D dynamic facial expression database. In *Proceedings of 8th IEEE International Conference on Automatic Face and Gesture Recognition*, pages 1–6, 17-19 Sept. 2008.