Data2Game: Towards an Integrated Demonstrator

Johannes Steinrücke[1], Paris Mavromoustakos-Blom[2], Judith van Stegeren[1], Ymko Attema[4], Sander Bakkes[3], Thomas de Groot[5], Johan de Heer[5], Dirk Heylen[1], Rafal Hrynkiewicz[5], Ton de Jong[1], Tije Oortwijn[5], Pieter Spronck[2], Mariët Theune[1], Bernard Veldkamp[1]

[1]University of Twente: {d.k.j.heylen, a.j.m.dejong, m.theune, j.e.vanstegeren, j.steinrucke, b.p.veldkamp}@utwente.nl; [2]Tilburg University: {p.mavromoustakosblom, p.spronck}@uvt.nl; [3]Utrecht University: s.c.j.bakkes@uu.nl; [4]Twente Safety Organization: y.attema@brandweertwente.nl; [5]Thales Research & Technology NL {thomas.degroot, johan.deheer, rafal.hrynkiewicz, tije.oortwijn}@nl.thalesgroup.com

Abstract. The Data2Game project investigates how the efficacy of computerized training games can be enhanced by tailoring training scenarios to the individual player. The research is centered around three research innovations: (1) techniques for the automated modelling of players' affective states, based on exhibited social signals, (2) techniques for the automated generation of in-game narratives tailored to the learning needs of the player, and (3) validated studies on the relation of the player behavior and game properties to learning performance. This paper describes the integration of the main results into a joint prototype.

Keywords: Serious games · Sensory data · Text generation · Player assessment

1   Introduction

The Data2Game project[1] investigates how, and to what extent, the efficacy of computerized training games can be enhanced by tailoring training scenarios to the individual player. The research is aimed at achieving three concrete research innovations: (1) techniques for the automated modelling of players' affective states based on exhibited social signals, such that the training scenarios can be tailored to the individual player, (2) techniques for the automated generation of in-game narratives that are tailored to the learning needs of the individual player, and (3) validated studies on the relation of the player behavior and game properties to learning performance. This paper reports on an intermediate version of the joint Data2Game prototype aiming to provide automated, personalized single player crisis management training sessions

2   Architecture and Building Blocks

For the joint Data2Game prototype the architectural diagram in Figure 2 served as a communication tool between the researchers / building block developers and the stakeholder of the DILEMMA game system. The logical component diagram led the

---

[1] https://www.nwo.nl/en/projects/05516114

development process and allowed for alignment and decision-making between all developers. The architecture includes four main components: a DILEMMA game (embedding a Training scenario), a Stress Model (based on Bio Sensory data), a Flavor Text Generator, and a Learning Assessment module (that connects via MQQT to R for statistical analyses). The goal of the Data2Game DILEMMA game is offering a microworld that enables studying both choice and decision-making behaviors. The goal of the Stress model is to provide a framework that does not only assess trainee stress during training, making it possible to adapt game components in real time to maintain trainee stress at an optimal level. The goal of the Flavor Text Generator is to make the in-game text adaptive to the player, by generating headlines that react on in-game player decisions. The goal of the learning assessment is to stimulate the player during game play to change a specific behavior such that a specific skill can be improved. Below we describe the main components in more detail.
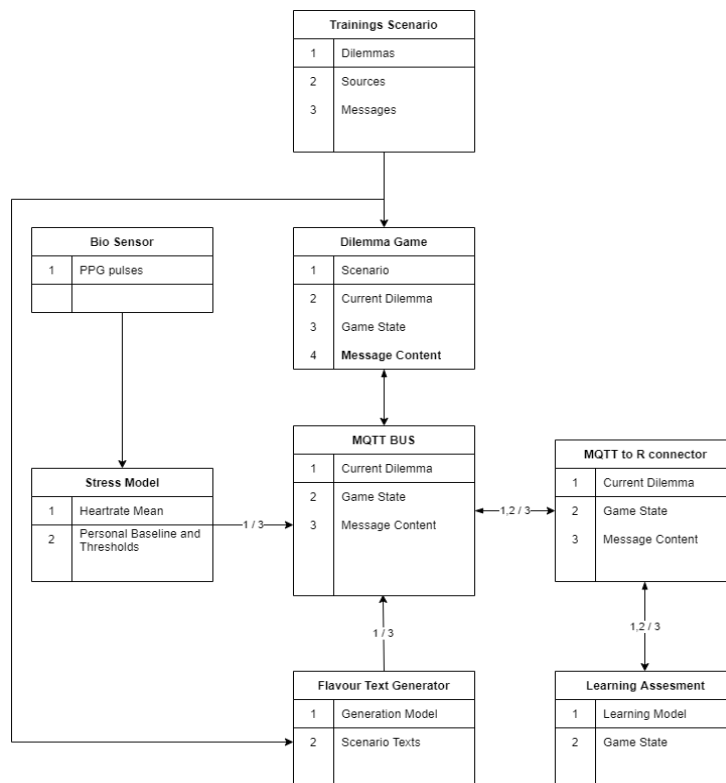

Fig 2. Architecture of the joint prototype

3   DILEMMA Game Engine and Data2Game Scenario

The DILEMMA [1] game engine was developed to elucidate human decision-making biases during choice behaviors as well as Dynamic Decision Making (DDM). A

DILEMMA game is considered a microworld that enables studying both choice and DDM behaviors [2]. It allows for the development of single player 2D (branching) narrative games, based around a decision room environment. During game play several dilemmas pop up that require a player's decision. Non-player Characters (NPCs) may provide information in the form of messages that could be relevant to the decision at hand. The player can request information from the NPCs or may even ask for advice regarding the required decisions. All player actions are logged and can be used for further analysis. The game scenario concerns a crisis management situation focusing on information literacy as the crucial skill in decision-making. The player has the role of the crisis manager and has final decision authority. Figure 3 (left) shows the interface of the game. NPCs provide information to a dilemma question that the player must answer.

Fig 3. Data2Game DILEMMA Game (left) and In-game expert feedback (right)

For the Data2Game project, in-game feedback is given by various experts: the stress specialist (based on the results of the Stress model), the text analysis specialist (based on the Flavor text generator module), and the learning specialist (based on the learning assessment research). Figure 3 (right) shows the graphical representation of in-game feedback given by the feedback specialists. The function of the in-game feedback is to encourage players to review their gameplay behavior in general, and reflect on specific aspects provided by the experts for further improvement.

4 Stress Model

We propose a framework for both measuring and manipulating crisis management trainee stress levels during training. The implementation of this framework has been divided into two parts: (1) processing and analyzing the physiological signals in real-time and (2) establishing communication between the sensor and game (Figure 4). To serve both purposes, an Express [11] back-end application was built on a local server, designed to receive, transmit and store game- and physiology-related data from and towards the game.

While the Shimmer3 GSR+ sensors are accompanied with the Consensys signal processing and analysis toolkit, we used a custom Python-based script for retrieving the raw physiological signals through a Bluetooth connection. Using the BiospPy library, the heart rate and skin conductance signals are processed, filtered, and analyzed. The BiospPy library is used to convert the raw heart rate signal into an analysis of the trainee's heart beats per minute (BPM). Similarly, the skin conductance signal is filtered

to remove artefacts from the signal. The analysis methods used in BiospPy are derived from Zong et al. [3] and Gamboa [4] for heart rate and skin conductance, respectively.
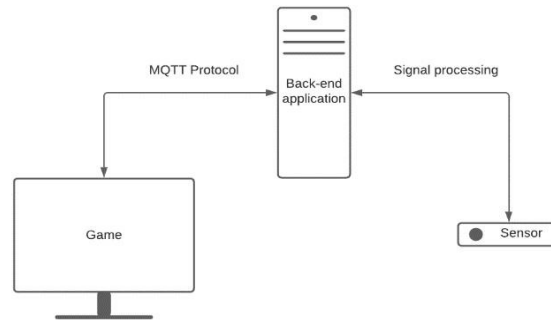

Fig 4. The architecture used to establish communication between the framework's components.

The processed heart rate and skin conductance signals are transmitted to the Express back-end application. Each data point transmitted is accompanied by an absolute timestamp, to ensure that the temporal nature of the signal is not harmed. All data points received by the back-end application are stored in memory until the recording session is finished and the back-end application is stopped. To establish a secure, persistent two-way connection between the game and the physiological sensors, we ensure that both sides are connected to the back-end application. To connect the game to the back-end application, an MQTT-based [5] connection was established. The connection is two-way, meaning that in-game actions can be serialized and sent to the back-end in the form of in-game events, but the back-end can also transmit messages to the game, in order to adapt in-game components.

At the start of the crisis scenario, the different components are started and connections between them are established. The sensors are attached to the trainee's non-dominant hand, and the raw signals are monitored and processed through the Python script. Necessarily, two-minute baseline physiological signals (in resting state) are extracted from the trainees before gameplay. Each processed physiological data point is transmitted to the back-end application as soon as processing is finished. When the trainee opens the description of the scenario's first dilemma, a start game message is transmitted from the game to the back-end application through MQTT, where the timestamp of the game's start will be stored. When trainees reach the expert feedback screen, a start feedback message is transmitted from the game to the back-end application. The back-end application will process the physiological data collected during gameplay and compare it to the previously collected baseline. An assessment of trainee stress levels will be extracted from the magnitude of difference between the baseline and gameplay heart rate signals, following the methods described in [6]. The assessment will be then transmitted to the game and shown to the trainees in the form of an expert feedback message.

5   Text Generation in the Crisis Game

Many role-playing boardgames and video games use flavor text, i.e., narrative elements that communicate the game's backstory, to draw the player into the game world and immerse them during play [7]. In this section we describe the text generator used by the Data2Game joint demo, which also creates flavor text in the form of headlines. The approach of underlying text generation system, called Churnalist, is to take existing headlines and adapt them to a new context: in this case, the context of the in-game dilemmas. Churnalist's architecture has been described in detail in [8] (Figure 5).
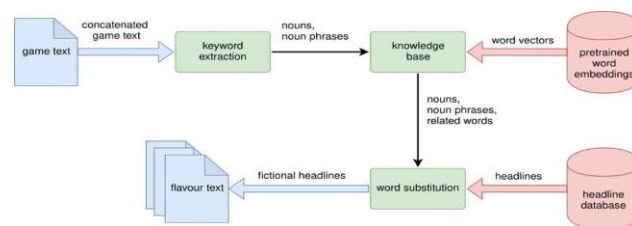

Fig 5. Churnalist system overview adapted from van Stegeren and Theune [8]

Churnalist analyses human-written crisis scenarios and dilemmas from the game for keywords. By reusing these words in a format that is familiar to the reader, i.e., headlines, we want to use the "charity of interpretation" effect to evoke a connection with the dilemmas in the player. The headline generator is built in Python 3, on top of a collection of open-source natural language processing libraries, such as NLTK [9], Pattern [10] and Spacy [11]. The module for the crisis game consists of the generator architecture, a dataset of headlines from the annotated Gigaword corpus [12], and a graphical user interface.

A game writer can use Churnalist to create new flavor text headlines for a given scenario. The first step is entering the human-written scenario in the text generation module. The generator uses text substitution, a light-weight language generation technique also used by poetry generators and Twitter bots. This method has advantages and disadvantages. Because the technique is not computationally expensive, the generator does not need a large source dataset to function, and because it is rule-based, it is 100% transparent. We can always inspect how the generator assembled a particular headline, and we can monitor the generator while it is running. However, the result of this low-cost approach is that the generated headlines are not always coherent at the sentence level. We mitigate this problem by asking the game writer to post-process the generated outputs. In the case of the Data2Game crisis game, which is in Dutch, approved headlines (in English) can be translated automatically with GoogleTranslate. Generating headlines first and translating them later is a low-cost method for text generation for relatively under resourced languages, such as Dutch.

The goal of the game's text generation module is to make the in-game text adaptive to the player, by providing flavor text headlines that react on in-game player decisions. Now, the text generator does not yet take player actions into account. Instead, the generated flavor text headlines are selected randomly during gameplay. However, the integrated system with player modeling and dynamic text presentation is the first step

towards a fully adaptive crisis game. An alternative approach to flavor text generation which we are currently investigating is using a neural language model based on Transformers [13], instead of the classic pipeline architecture for NLG systems [14]. It is possible to fine-tune a pretrained Transformer-based language model, such as GPT-2 [15] or BERT [16], on annotated headlines, and use this model as a text generator with comparable functionality to that of Churnalist. Training the neural model can take a long time – a few hours – and requires a much larger source dataset than is needed by Churnalist. However, once the language model has completed its training, it can generate headlines much faster and possibly with higher output quality. GPT2 [15], a pre trained language model that is also based on the Transformers architecture, performs particularly well for generation tasks. The drawback of a neural language model is that it is not transparent. We cannot inspect the language model in detail, and we cannot trace the origin of generated headlines. It also requires more computing power to perform the training and generation phase.

6   Player Assessment and Instructional Interventions

A crucial skill in crisis management decision-making is information literacy (IL). As a multifaceted skill, it highlights the importance of the decision-maker's ability to utilize their own experiences and the contextual information in the Data2Game crisis management decision-making scenario [17]. While purely relying on experience might lead to suboptimal decisions and choices made [18], completely analyzing and thoroughly weighing off all information is not desirable also, given that crisis situations can become quite complex in a rather short time span [19]. Summarizing, decision-makers cannot rely on their own experience only, but they also cannot analyze all contextual information without worrying about time constraints.

As stated, IL is a multifaceted skill: It can be defined as the ability to recognize the need of information (Standard 1), to locate information (Standard 2), to evaluate information and its sources (Standard 3), to interpret and to use the information (Standard 4); [20]. In their research, Steinrücke, Veldkamp [17] developed an unobtrusive method of assessing the players IL performance using the Data2Game Dilemma Game, by classifying the players' in-game behavior into IL skills classes (see Fig 6) using latent profile analyses. Despite being reliant on a self-report measure, this approach is considered promising [21].

Based on their gameplay behavior, the players can be placed in one of three, respectively four, classes, where a higher class indicates a higher skill level. Using a feedback moment, the game could then stimulate the player to change a specific behavior such that the next highest class can be reached. For example, if the player is classified in class three out of four on Standard 3, the game would prompt the player to re-evaluate already evaluated information more often. This feedback moment could be supported by a reflective moment, which could promote the players to review their gameplay behavior in general, and critically reflect on possible points of improvement [22]. Currently, the effect of a feedback moment, incorporating a self-reflection and a social comparison [23, 24], on the gameplay behavior and on the in-game IL performance is being investigated [25].
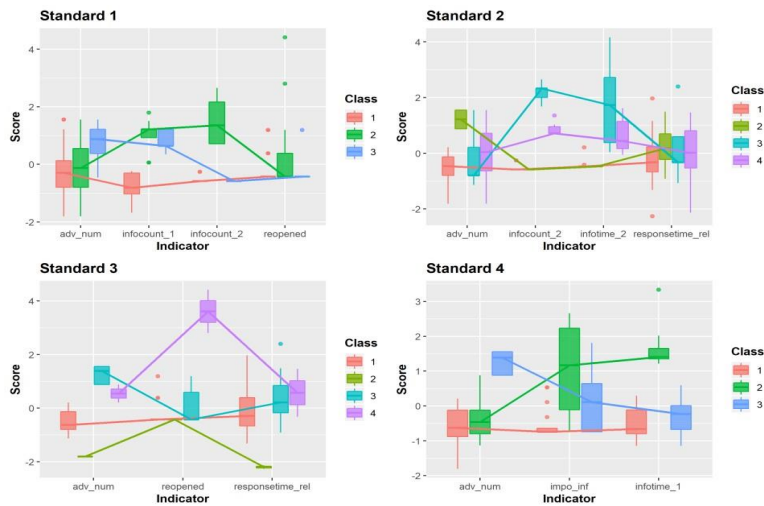
Fig 6. Gameplay profiles for the different facets of IL. Adapted from [17].

7 Conclusion

This paper presented an intermediate version of the joint Data2Game prototype aiming to provide fully automated, personalized single player crisis management training sessions. The four main components were described, including next steps regarding several research directions. In sum, (1) based on the stress model we shall adapt the game components in real time to maintain stress at an optimal level, (2) the text generator shall take player actions into account, and (3) the learning assessment feedback shall incorporate a self-reflection takes a social comparison into account.

References

1. de Heer, J., de Groot, T., Hrynkiewicz, R., Oortwijn, T., & Porskamp, P. Devilish Dilemma Games: Narrative Game Engine and Accompanying Authoring Tool. In: AHFE 2019, pp. 169-178. Springer, Cham (2019)
2. Gonzalez, C., Lerch, J.F., Lebiere, C. Instance-based Learning in Dynamic Decision Making. Cognitive Science 27, 591--635 (2003)
3. Zong, W., Heldt, T., Moody, G. and Mark, R. An Open-Source Algorithm to Detect Onset Arterial Blood Pressure Pulses. In: Computers in Cardiology, 2003, pp. 259–262, IEEE (2003)
4. Gamboa, H. Multi-modal Behavioral Biometrics Based on HCI and Electrophysiology. PhD Thesis, Universidade Técnica de Lisboa, (2008)

5. Light, R. A. Mosquitto: Server and Client Implementation of the MQTT Protocol. Journal of Open Source Software, 2(13), p. 265 (2017)
6. Villarejo, M. V., Zapirain, B. G. and Zorrilla, A. M. A Stress Sensor Based on Galvanic Skin Response (GSR) Controlled by Zigbee. Sensors, 12(5), 6075--6101 (2012)
7. Bateman. C. Game Writing: Narrative skills for Videogames. Charles River Media (2007)
8. van Stegeren, J. and Theune, M. Churnalist: Fictional Headline Generation for Context appropriate Flavor Text. In: Proceedings of the Tenth International Conference on Computational Creativity. ICCC '19, pp. 65–72. Charlotte, NC, USA: Association for Computational Creativity (ACC) (2019)
9. Bird, S., Klein, E. and Loper, E. Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. O'Reilly Media, Inc. (2009)
10. De Smedt, T. and Daelemans, W. Pattern for Python. The Journal of Machine Learning Research, 13(1), 2063--2067 (2012)
11. Honnibal, M. and Montani, I. SpaCy 2: Natural Language Understanding with Bloom Embeddings, Convolutional Neural Networks and Incremental Parsing. To appear (2017)
12. Napoles, C., Gormley, M. and Van Durme, B. Annotated GigaWord. In: Proceedings of the Joint Workshop on Automatic Knowledge Base Construction and Web-scale Knowledge Extraction, pp. 95–100. Association for Computational Linguistics (2012)
13. Vaswani A. et al. Attention is all you need. In: 31st Conference on Neural Information Processing Systems (NIPS 2017), pp. 5998–6008 (2017)
14. Reiter, E. and Dale, R. Building Applied Natural Language Generation Systems. Natural Language Engineering, 3(1), 57--87 (1997)
15. Radford, A., et al. Language Models Are Unsupervised Multitask Learners. In: https://github.com/openai/gpt-2. Retrieved August 31, 2020 (2019)
16. Devlin, J., et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics, pp. 4171–4186. Association for Computational Linguistics (2019)
17. Steinrücke, J., Veldkamp, B.P., and de Jong, T. Information Literacy Skills Assessment in Digital Crisis Management Training for the Safety Domain: Developing an Unobtrusive Method. Frontiers in Education, 5(140) (2020)
18. Mezey, G. Crisis Management Decision Making. AARMS, 3, 267-288 (2004)
19. van der Hulst, A.H., et al. Serious Gaming for Complex Decision Making: Training Approaches. Int. Journal of Technology Enhanced Learning, 6(3), 249--264 (2014)
20. Information Literacy Competency Standards for Higher Education. 2000, https://alair.ala.org/bitstream/handle/11213/7668/ACRL%20Information%20Literacy%20Competency%20Standards%20for%20Higher%20Education.pdf
21. Hahnel, C., Eichmann, B., and Goldhammer, F. Evaluation of Online Information in University Students: Development and Scaling of the Screening Instrument EVON. Frontiers in Psychology, 11(3460) (2020)
22. Wouters, P., and van Oostendorp, H.: A Meta-Analytic Review of the Role of Instructional Support in Game-based Learning. Computers & Education. 60(1), 412--425 (2013)
23. Davis, D., Jivet, I., Kizilcec, R. F., Chen, G., Hauff, C., Houben, G.-J.: Follow the Successful Crowd: Raising MOOC Completion Rates through Social Comparison at Scale. In: Proceedings of the 7th Int. Learning Analytics & Knowledge Conference (2017)
24. Jivet, I., Scheffel, M., Drachsler, H., and Specht, M. Awareness Is Not Enough: Pitfalls of Learning Analytics Dashboards in the Educational Practice. In É. Lavoué, H. Drachsler, K. Verbert, J. Broisin & M. Pérez-Sanagustín (Eds.), Data Driven Approaches in Digital Education. LNCS, vol. 10474, pp. 82-96. Springer, Heidelberg (2017)
25. Steinrücke, J., Veldkamp, B.P., and de Jong, T. The Effect of Self-Reflection on Information Usage and Information Literacy in a Digital Serious Game (in preparation)