3.13 Explainable AI for Games

Jichen Zhu (IT University of Copenhagen, DK), Maren Awiszus (Leibniz Universität Hannover, DE), Michael Cook (Queen Mary University of London, GB), Alexander Dockhorn (Leibniz Universität Hannover, DE), Manuel Eberhardinger (Hochschule der Medien – Stuttgart, DE), Daniele Loiacono (Polytechnic University of Milan, IT), Simon M. Lucas (Queen Mary University of London, GB), Ana Matran-Fernandez (University of Essex – Colchester, GB), Diego Perez Liebana (Queen Mary University of London, GB), Tommy Thompson (AI and Games – London, GB), and Remco Veltkamp (Utrecht University, NL)

In recent years more and more research has been invested into eXplainable artificial intelligence (XAI) to make machine learning (ML) and AI models more trustworthy and understandable for users. In an earlier vision paper, a new research area for designers and game designers was proposed called XAI for Designers (XAID) [1], which focused on *mixed-initiative co-creation* [2] approaches to help designers better leverage AI methods through co-creation in their respective design tasks. Since then, much development has been made in XAI. In this working group, we investigate whether and how these new methods for XAI can also be used for games.

3.13.1 What is XAI for Games?

There are a large variety of possible use cases for XAI in games or game development, and this largely depends on what one wants to achieve. Some salient use cases include:

- Increasing the transparency for game AI decisions so that these decisions can be understood and trusted by humans.
- Explanations of key game AI decisions can be used as a feedback mechanism for how well a player is performing. For instance, a PCG-based educational game can explain to a player that a new level is generated based on her previous gameplay so that she can continue to practice a certain skill that she has not mastered. This type of explanation can be used as a feedback mechanism to foster player reflection and learning[6].
- Tools for framing in computational creativity and improving the design experience with mixed-initiative co-creativity systems.
- Highlighting to players why a given strategy is relevant, optimal, or exciting.

To further narrow the focus on the different use cases, in this report, we will focus on procedural content generation with ML (PCGML).

3.13.2 Case Study: Mario Level Generation

First, we looked at different possibilities to generate Super Mario levels. TOAD-GAN [3] can be trained using only one example. This method also makes it possible for users to control the output of the generation process by changing the noise vector that represents the input of the generator network. Since noise vectors cannot be interpreted by designers, designers still do not have the ability to design content according to their needs. To accomplish this, one must make the noise vector explainable to designers and map the different areas of the noise vector to the content that would result from a change in the noise vector.

Another method for generating Super Mario levels uses an evolutionary algorithm with tilesets [4]. The tilesets enforce consistency of the output, and the Kullback-Leiber Divergence

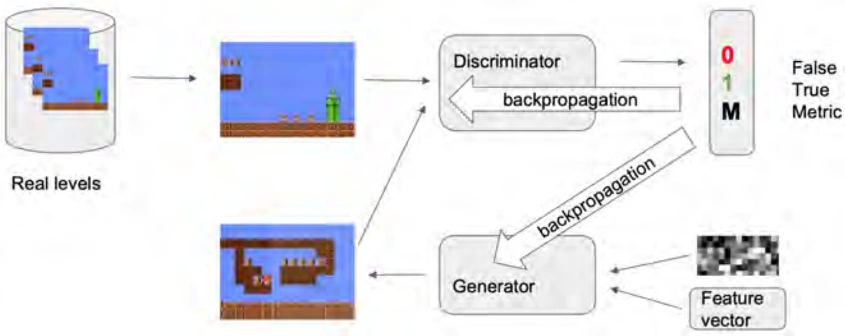

Figure 9 An overview of DOOMGAN architecture.

enables for control of variation and novelty. This method is explainable by design as the history of the gene values and the time steps of the mutation operators could be used to identify *when* something occurred, and *why* it was picked to be modified.

3.13.3 Case Study: DOOMGAN – Improving the PCGML Interpretability by Incorporating Metrics

PCGML[5] has been successfully applied to several kinds of game content. However, it generally has low interpretability to human designers because how the input (e.g., parameter/feature vectors) leads to generated content (or corresponding gameplay metrics) is often opaque. Recent Deep Learning-based generative models exacerbate this problem due to their complexity and blackbox nature. As a relevant case of study, we focused on GAN-based PCGML approaches and proposed to incorporate gameplay metrics (e.g., completion time, win rate) in part of the GAN architecture at the level of the discriminator. Figure 9 provides an overview of the proposed GAN architecture, dubbed DOOMGAN, where the discriminator is extended by adding one or more gameplay metrics as additional outputs. Our research hypotheses are that this method will 1) improve the interpretability of the system by providing meaningful intermediate output to designers and 2) improve the performance of the generative model (e.g., better data quality and data efficiency). Moreover, with the proposed method, existing XAI techniques, such as Saliency Map, LIME, and DeepSHAP, can be used to further open the blackbox of PCGML. An ideal testbed to investigate our ideas would be to extend one of the Mario level generators based on GAN previously introduced in the literature (e.g., TOAD-GAN[3]). To the best of our knowledge, this is among the first approach that connects XAI to PCGML methods.

3.13.4 Open Problems

Explainable AI for games is still a nascent research area. Below we summarize some of the key open problems in this area:

- How to turn explainability into explanation and actionable explanations to players and/or designers?
- How does content representation affect explainability? (e.g., representing a Mario level as tiles vs. objects)
- Whom do we design the explainable system for? What do the human players, designers, or other stakeholders need? Current XAI methods only explain predictive models but not generative models.
- How to capture functionality/playability of a level in XAI, which is absent in image generation?

3.13.5 Conclusion

In summary, this working group found eXplainable AI to be a rich research topic to explore in the context of computer games. Making the underlying AI process more transparent can benefit a wide range of stakeholders, including players, game designers, game analytics/user researchers, and game producers. Since computer games are end user-facing, we believe exploring eXplainable AI in the context of games will expedite the transition from technical explainability to usable human-centered explanations.

References

1 Zhu, J., Liapis, A., Risi, S., Bidarra, R. & Youngblood, G. Explainable AI for Designers: A Human-Centered Perspective on Mixed-Initiative Co-Creation. *2018 IEEE Conference On Computational Intelligence And Games (CIG)*. pp. 1-8 (2018)

2 Yannakakis, G., Liapis, A. & Alexopoulos, C. Mixed-initiative co-creativity. *FDG*. (2014)

3 Awiszus, M., Schubert, F. & Rosenhahn, B. TOAD-GAN: Coherent Style Level Generation from a Single Example. *AAAI Conference On Artificial Intelligence And Interactive Digital Entertainment Best Student Paper Award* . (2020,10)

4 Lucas, S. & Volz, V. Tile Pattern KL-Divergence for Analysing and Evolving Game Levels. *Proceedings Of The Genetic And Evolutionary Computation Conference*. pp. 170-178 (2019), https://doi.org/10.1145/3321707.3321781

5 Summerville, A., Snodgrass, S., Guzdial, M., Holmgård, C., Hoover, A., Isaksen, A., Nealen, A. & Togelius, J. Procedural Content Generation via Machine Learning (PCGML). *IEEE Transactions On Games*. **10**, 257-270 (2018)

6 Zhu, J. & El-Nasr, M. Open player modeling: Empowering players through data transparency. *ArXiv Preprint ArXiv:2110.05810*. (2021)

3.14 Human-AI Collaboration Through Play

Jichen Zhu (IT University of Copenhagen, DK), Guillaume Chanel (University of Geneva, CH), Michael Cook (Queen Mary University of London, GB), Alena Denisova (University of York, GB), Casper Harteveld (Northeastern University – Boston, US), and Mike Preuß (Leiden University, NL)

3.14.1 Motivation

Human-AI collaboration is a rapidly growing research area. As AI becomes an integral part of the workplace as well as home, developing technology that can efficiently collaborate with humans is essential.

Existing psychology research found that successful collaborations between humans need the foundation of 1) a relational interaction (conflict, small talk, emotional exchanges, relationship construction) and 2) efficient cognitive interaction (e.g., building on others' ideas – transactivity, synthesis, building a common ground) [1]. However, in current *Human-AI interaction* (HAI) research, this social-cognitive element and the social experience between human users and the AI is under-explored. This is problematic because since most users, especially novel users of AI, tend to approach AI based on their knowledge of similar human interactions [9, 8].