Cybersickness Analysis Using Symbolic Machine Learning Algorithms

Thiago Porcino¹, Daniela Trevisan¹, Esteban Clua¹

¹ Instituto de Computação – Universidade Federal Fluminense

{thiagomp,daniela,esteban}@ic.uff.br

Abstract. Virtual reality (VR) and head-mounted displays are constantly gaining popularity in various fields such as education, military, entertainment, and health. Although such technologies provide a high sense of immersion, they can also trigger symptoms of discomfort. This condition is called cybersickness (CS) and is quite popular in recent VR publications. This work proposes a novel experimental analysis using symbolic machine learning to rank potential causes of CS in VR games. We estimate CS causes and rank them according to their impact on the classical machine learning classification task. Experiments are performed using two VR games and 6 experimental protocols along with 37 valid samples from a total of 88 volunteers.

1. Introduction

Activities such as virtual training environments, simulations and entertainment in immersive virtual formats are constantly becoming more popular with the continued development and public interest in VR technologies over the last years [Calvelo et al. 2020]. In 2019, the VR hardware market was valued at 4.4 billion US dollars and is expected to reach 10 billion US dollars by 2022 [Statista 2020].

Head-mounted displays (HMDs) is one of the means of achieving immersive virtual reality. These devices usually consist of electronic displays and lenses that are fixed over the head where the display and lenses face the eyes of the user. HMDs are used for various purposes in the industry such as in games that focus on entertainment [Studios 2015], military [Rizzo et al. 2011], education [Ahir et al. 2020], therapy [Carrión et al. 2019] and simulators for numerous contexts [Kühnapfel et al. 2000].

Unfortunately, HMDs are strongly related to frequent manifestations of discomfort [Kolasinski 1995]. Among the possible manifestations, cybersickness (CS) deserves special attention as it is the most frequent and is usually associated to long exposures to HMDs. According to Ramsey et al. [Ramsey et al. 1999], approximately 80% of participants who have already experienced HMD-based VR reported discomfort sensations after just 10 minutes of exposure. In addition, more than 60% of usability problems are strongly related to discomfort [Kolasinski 1995].

The most frequent symptoms caused by CS are general discomfort, headache, stomach awareness, nausea, vomiting, sweating, fatigue, drowsiness, disorientation, and apathy [Dennison and D’Zmura 2017]. These symptoms impact the user experience and affect the profit and coverage of the VR game industry. In addition, discomfort symptoms can vary across individuals, where some people are more susceptible than others.
Several works in the literature address the CS phenomenon and mitigation strategies for immersive VR applications using HMDs [Mousavi et al. 2013, Davis et al. 2014, Rebenitsch and Owen 2016, Porcino et al. 2020b]. While most previous works [Jin et al. 2018, Kim et al. 2019, Jeong et al. 2019] are mainly focused on detecting and predicting CS events, this work estimates and is amenable to rank the attributes that contribute the most to triggering cybersickness, enabling a selection of the most adequate strategy to mitigate CS. We propose an approach that allows the estimation of a cause while the user is under the VR condition. However, the suggestion of strategies can be implemented afterwards by the game designer. In this direction, symbolic machine learning is adequate, as understandability is essential [Maree and Omlin 2020].

2. Contributions and Awards

Through this research, we have been contributing to mitigate cybersickness problems in VR environments:

- **Providing an extensive cybersickness literature review.** Our review facilitates researchers to identify the leading causes for most discomfort situations in VR environments and associate the most recommended strategies to minimize such discomfort. Additionally, Kemeny et al. [Kemeny et al. 2020] cited part of this work in their book. At the moment, more than 90 works (not only in computing, but also in other areas such as medicine and psychology) cited us [Porcino et al. 2017].

- **SVR’ best review paper award.** Part of this work [Porcino et al. 2020b] was awarded at the symposium on virtual and augmented reality (SVR), in 2020. Additionally, we published an extended version of this study to SBC’ Journal on Interactive Systems [Porcino et al. 2021d].

- **Proposing the cybersickness profile questionnaire.** We create the cybersickness profile questionnaire (CSPQ) based on our literature findings. The CSPQ contains 9 questions about user profile tied to cybersickness manifestations [Porcino et al. 2020a].

- **Proposing symbolic machine learning models to identify causes of cybersickness in VR environments.** We are the first to use symbolic classifiers (decision tree and random forest) to estimate CS causes during a gameplay experience [Porcino et al. 2021b]. Additionally, we published an extended version of this study to Entertainment Computing Journal [Porcino et al. 2021a].

- **Proposing an experimental methodology to capture user and gameplay data tied to cybersickness.** We created and conducted an iterative evaluating protocol methodology and proposed two VR games (a racing game and a flight game) for user and gameplay data acquisition [Porcino et al. 2021e].

- **SVR’ best doctoral thesis award.** This work was awarded at the symposium on virtual and augmented reality (SVR), in 2021 [Porcino et al. 2021c].

- **Providing a public VR users database.** The raw dataset of this work is published in a public domain for further reproduction and comparisons [Porcino 2021].

2.1. Related Work

Several studies use deep learning models such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to predict CS. Kim, J. et al. [Kim et al. 2019],
proposed a deep learning architecture to estimate the cognitive state using brain signals and how they are related to CS levels. Their approach is based on deep learning models, such as long short-term memory (LSTM) RNN and CNN [Lawrence et al. 1997, Graves et al. 2013, Sak et al. 2014]). The algorithms learn the individual characteristics of the participants that lead to the manifestation of CS symptoms when watching a VR video.

Jin et al. [Jin et al. 2018] grouped CS causes as follows: hardware characteristics (VR device settings and features), software characteristics (content of the VR scenes), and characteristics based on the individual user. The authors used classifiers to estimate the level of discomfort. A total of three machine learning algorithms (CNN, LSTM-RNN, and support vector regression [Drucker et al. 1997]) were used. According to the authors, the LSTM-RNN obtained the best results.

Jeong et al. [Jeong et al. 2019] focused on 360° VR streaming content. They analyzed scenarios of CS using brain signals. Their work uses data from 24 individuals and aims to identify common brain characteristics of VR stream patterns associated to CS manifestation. They examined VR segments and highlighted them when several individuals felt discomfort at the same time. However, the authors were not able to find specific and individual CS causes.

Garcia-Agundez et al. [Garcia-Agundez et al. 2019] focus on the classification of the level of CS. The proposed model uses a combination of bio-signal and game settings. User signals such as respiratory and skin conductivity of 66 participants were collected. Authors obtained a classification accuracy of 82% (SVM) when it comes to binary classifications and 56% (kNN) for the ternary case. Moreover, Porcino et al. [Porcino et al. 2020a] proposed an approach that is amenable to the prediction of CS during the gameplay. Authors were able to achieve an average accuracy of 96.54% with random forest considering a total of 16 different machine learning algorithms and different scenarios. In addition, they identified attributes responsible for painful states in VR games.

3. Proposed Solution

Kim, J. et al. [Kim et al. 2019] and Jeong et al. [Jeong et al. 2019] capture data using external medical equipment. This equipment is not mainstream in terms of VR. In this work, we focus on data captured without the use of extra accessories. Hence, we discard the use of any external medical equipment that could harm the user experience, or even that decrease the likelihood of the user owning the equipment in the first place.

Furthermore, Garcia-Agundez et al. [Garcia-Agundez et al. 2019] and Porcino et al. [Porcino et al. 2020a] do not focus on the estimate of the weight or on the influence of the attributes (i.e., the cause) leading to CS, as opposed to the proposal of this work.

In this novel approach, we use symbolic machine learning to analyse and identify one or more causes of discomfort, which is user and context specific. In other words, the approach described in this manuscript is not a general rule for recognizing the presence of discomfort as previously approached in the current literature. In contrast, it provides real-time user and context-sensitive evaluation and estimation of causes for cybersickness.

Moreover, the use of symbolic classifiers is paramount for an appropriate anal-
ysis and understanding of the decision. Although previous work suggest that deep learning classifiers are the most suitable approach for CS prediction [Jeong et al. 2019, Kim et al. 2019], deep neural networks are black boxes that are very difficult to grasp. For this reason, this research limited the analysis to symbolic machine learning algorithms that enable a straight understanding of the decision path.

This work is built upon the understandability of symbolic classifiers, whose prediction process can be represented by a set of unordered or disjoint rules. Symbolic classifiers are not novel and they have been used in many scenarios where clear logical comprehensiveness is required [Bernadini 2006]. Among the comprehensible realm of classifiers we can highlight decision tables, bayes classifiers and most notably in terms of accuracy rates, decision tree approaches. Decision tree approaches offer a robust prediction and a wide variety of algorithms and implementations in the literature [Rodrigues et al. 2018]. A decision tree can be written as a set of disjoint unordered rules [Flach 2012].

The logical prediction path of the decision trees inherit a personal fingerprint associated to attribute weights. Usually, attributes that are closer to the tree root are more important, as they often reduce the chaos in data more than the rest, i.e., they separate the information more appropriately than other features, improving the information gain and reducing entropy. As a general rule, the frequency in which attributes appear in the decision path is also an important piece of information. We combine these two aspects to estimate the importance of the attribute (i.e., the most important causes of discomfort).

Let us suppose a decision tree described by 13 decision nodes, as shown in Figure 1, which contains in their conditions the features Gender, Rotation, VR experience, and Acceleration. Furthermore, let us consider a path for an instance that was predicted as discomfort, where the green arrows illustrate a single decision path.

The coverage of each node is also shown in Figure 1. We can observe that the lower the depth of the three, the lower the coverage of each node. We used this aspect to calculate the importance of the features. In this sense, we compute a potential-cause score (PCS) by summing up the heights of these features (e.g., for a specific instance, as highlighted with the green arrows). In this case, gender appeared once and has height 3, VR experience has height 2, and Rotation has height 1. Next, the output is divided by the sum of all depths of the tree, as follows: gender=3/6 or 50, VR experience=2/6 or approximately 33 and Rotation=1/6 or approximately 17. In this case, we estimate gender as the most relevant cause for CS.

\[
P_{\text{CS}}(F) = \frac{\sum_{h=0}^{\text{max height}} \begin{cases} h, & \text{if } F \text{ belongs to the height}\ 0, \\ 0, & \text{otherwise} \end{cases}}{\sum_{h=0}^{\text{max height}} h} \tag{1}
\]

In equation 1, \( h \) varies from 0 to the maximal tree height, and \( F \) is the feature being evaluated. PCS is computed considering just the decision path (e.g., the one highlighted in green).

Furthermore, the random forest model can be considered a set of decision trees. We sum the PCS results from each tree \( t \) if this tree’s final decision is equal to the RF final decision. Otherwise, we sum 0 in this iteration. In Equation 2, we sum the PCS results from each tree \( t \) if the classification result of \( t \) is equal to the final RF classification result,
where votes from several trees are scored together and a single class, e.g., the mode, is chosen as the classification result. Otherwise, we sum 0 at this iteration.

$$PCS_{RF}(F) = \sum_{t \in T} \{ PCS(F), \text{if } tree\ t\ decision = RF\ final\ decision 0,\ otherwise \}$$

(2)

4. Results

Predicting the cause of CS is not trivial. Every user has a specific susceptibility to discomfort. Furthermore, several attributes are related to the hardware and ergonomic aspects of the devices. We are still far from tracing very precise causes for all specific cases. However, so far, factors such as rotation, speed, gender, and previous VR experience, appeared as dominant factors that can trigger CS.

Our approach works with up to eight factors attributed to CS. Previous works in the literature already proposed strategies for four of these attributes (exposure time, acceleration, speed, frame rate, and camera rotation on the z-axis) [Melo et al. 2018, Bouyer et al. 2017, Budhiraja et al. 2017, Van Waveren 2016]. The remaining causes (gender, VR experience, and age) are causes associated to the user profile and are still not associated to a clear strategy. In addition, we observed different patterns of causes for users in the race game when compared to the flight game.

Exposure time (timestamp) was the most frequent cause of discomfort. Overall, the race game contributed more to CS manifestation (39.4) when compared to the flight game (35.9). A possible suggestion is to reduce the time of exposure in the case of race games. CS triggered by acceleration shifts controlled by users was less frequent in the race game (5.6) when compared to the flight game, where acceleration was not controlled by the user (11.80).

Another feature that influences on the discomfort in both scenarios is the former VR experience. PCS was greater in the race game in contrast to the flight game, 8.25 and
Figure 2. Random Forest feature ranking (identification of cybersickness causes) for the race (A) and flight game (B) for P5 subjects.

4.76, respectively. In addition, rotation was marked as cause more frequently in the flight (18.70) when compared to the race game (13.83).

Our results (Figure 2) show that rotation and acceleration triggered cybersickness more frequently in a flight game in contrast to a race game. We could also observe that participants that are less experienced with VR are more prone to feel discomfort. Former experience plays a more important role on the race game, as this game provides more liberty to the user in terms of controllers, more displacement alternatives and a more self-controlled acceleration. Additionally, time is a crucial variable in terms of CS mitigation. Conclusively, different causes that trigger discomfort arise based on short or long term VR exposures. As final remark, we suggest strategies for mitigating CS for these two scenarios.

5. Conclusions

In this work, we propose an approach to identify causes of cybersickness in different virtual reality games using head-mounted displays. To the best of our knowledge, this is the first work that uses symbolic classifiers to analyze causes of cybersickness during the gameplay experience. Once the cause is identified, game designers are able to select the most adequate strategy to mitigate the impacts of CS, according to the literature.

We experimented with two different scenarios and proposed the use of two symbolic machine learning algorithms, along with an analysis to identify the optimal tree depth for the generated models. Next, we performed a feature ranking to identify the most relevant causes of CS, which vary along with the gameplay experience and is related to the user. We observed that exposure time, rotation, and acceleration are most likely the top factors contributing to CS. As exposure time appeared as a preliminary cause for CS in our experiments, we suggest reducing the time of exposure in the case of race games (to 5 minutes max) or to provide intervals for the user at every 5 minutes of experience.

At last, we conclude that introducing rapid movements and related variables that are controlled by the user can potentially lead to higher incidence of cybersickness. Virtual reality games that rely on low complexity controllers are a better fit for non-experienced users.
As a final remark, the raw dataset of this work along with developed games are available in the following public domain: for further reproduction and comparison [Porcino 2021].

5.1. Limitations and discussion

COVID-19 pandemic affected our experiments and protocols in terms of dataset construction. For this reason, some features were also not well represented, such as gender (for women), age (for older adults), and experience (for people with former VR experience). Moreover, the number of used games did not cover locomotion movements, which is specific for games where the user can walk virtually.

Another concern is related to a lack of a more robust treatment of the CS level defined by the verbal feedback of the participants. In other words, verbal feedback is highly subjective and degrees vary from each participant (e.g., moderate for a user may not be moderate for another). Consequently, it is tough to define a robust CS feedback analysis. For this reason, we also considered a binary feedback: no discomfort (for 0 levels of discomfort) and discomfort (for levels 1, 2, and 3).

5.2. Future work

Future work involves including features such as posture, vision impairments, locomotion and other information into our framework. We also aim to improve the balance of the dataset as some cases were also not broadly represented, such as gender (few women), age (few elders) and experience (few subjects with former VR experience). Moreover, as Exposure time was observed as the most frequent cause of CS in our experiments, we also intend to conduct investigative research to detect in which specific timestamp we can identify the increased manifestation of CS symptoms.

Another straightforward way is to explore the gender differences tied to games and virtual reality tasks. Our results and other works [Liang et al. 2019, Grassini and Laumann 2020, Curry et al. 2020] pointed out that specific tasks can produce different results of discomfort for different user profiles and groups, regarding and not limited to: gender, age, or health issues. Our symbolic machine learning approach can also assist with future analyses.

Moreover, it is necessary to better understand the correlation of profile features with gameplay features, and also how results obtained from profile features (gender, age, VR experience) can be used to label VR experiences according to different groups of users.

Another challenge is to create a virtual reality experience that explores specific tasks individually tied to specific CS causes with a long exposure. The evaluation of individual tasks associated with CS causes may produce a more profound study isolating any other VR potential influences on the CS results.

At last, we would like to highlight that this work can be seen as a preliminary guide to elaborate more adequate game designs for VR games and related applications.
References


in virtual reality games. In *2020 IEEE 8th International Conference on Serious Games and Applications for Health (SeGAH)*, pages 1–8. IEEE.


