Expert Systems with Applications 41 (2014) 7281–7290

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Artificial Intelligence approaches for the generation and assessment of believable human-like behaviour in virtual characters

Joan Marc Llargues Asensio [a], Juan Peralta [a], Raul Arrabales [b], Manuel Gonzalez Bedia [c,*], Paulo Cortez [d], Antonio Lopez Peña [a]

[a] Universidad Autnoma de Barcelona, Edifici O, Campus UAB 08193 Bellaterra, Barcelona, Spain
[b] Conscious Machines Lab, Madrid, Spain
[c] Universidad de Zaragoza, Calle Pedro Cerbuna, 12, 50009 Zaragoza, Spain
[d] ALGORITMI Research Centre, University of Minho, Guimaraes, Portugal

ARTICLE INFO

Article history:
Available online 23 May 2014

Keywords:
Turing test
Human-like behaviour
Believability
Non-player characters
Cognitive architectures
Genetic algorithm
Artificial neural networks

ABSTRACT

Having artificial agents to autonomously produce human-like behaviour is one of the most ambitious original goals of Artificial Intelligence (AI) and remains an open problem nowadays. The imitation game originally proposed by Turing constitute a very effective method to prove the indistinguishability of an artificial agent. The behaviour of an agent is said to be indistinguishable from that of a human when observers (the so-called judges in the Turing test) cannot tell apart humans and non-human agents. Different environments, testing protocols, scopes and problem domains can be established to develop limited versions or variants of the original Turing test. In this paper we use a specific version of the Turing test, based on the international BotPrize competition, built in a First-Person Shooter video game, where both human players and non-player characters interact in complex virtual environments. Based on our past experience both in the BotPrize competition and other robotics and computer game AI applications we have developed three new more advanced controllers for believable agents: two based on a combination of the CERA–CRANIUM and SOAR cognitive architectures and other based on ADANN, a system for the automatic evolution and adaptation of artificial neural networks. These two new agents have been put to the test jointly with CCBot3, the winner of BotPrize 2010 competition (Arrabales et al., 2012), and have showed a significant improvement in the humanness ratio. Additionally, we have confronted all these bots to both First-person believability assessment (BotPrize original judging protocol) and Third-person believability assessment, demonstrating that the active involvement of the judge has a great impact in the recognition of human-like behaviour.

1. Introduction

The design and implementation of believable artificial agents, truly indistinguishable from humans, remains an open problem. This challenge has been typically addressed from two interrelated perspectives within cognitive science. On one hand, psychological models of human cognition try to explain how human behaviour is produced. On the other hand, computational models implemented in artificial agents try to replicate to some extent human-like behaviour. In this work, we focus exclusively in the sensorimotor behavioural dimension, setting aside any concerns related to the physical appearance of the artificial agents or their verbal report capabilities.

The imitation game proposed by Turing is the paradigmatic test for believability. However, current state of the art in cognitive and computer sciences has not reached the degree of development in which this test could be considered truly achievable. Therefore, a number of different variations of the original Turing test have been proposed, usually limited Turing tests with relaxed constraints and more specific problem domains. In this paper, we focus in a specific limited version of the Turing test designed for virtual characters and based in a First-Person Shooter (FPS) video game.

From the point of view of the scientific research on human cognition, video game characters are an interesting case of artificial agents because they are easy to implement using the video game industry state of the art tools and their virtual environments can become quite complex, simulating a great variety of contexts and

* Corresponding author. Tel.: +34 93581 4090.
E-mail addresses: fireball787b@gmail.com (J.M. Llargues Asensio), jperalta@cvc.uab.es (J. Peralta), raul@conscious-robots.com (R. Arrabales), mgbedia@unizar.es (M.G. Bedia), pcortez@dsi.uminho.pt (P. Cortez), antonio@cvc.uab.es (A.L. Peña).

http://dx.doi.org/10.1016/j.eswa.2014.05.004
0957-4174/

ambient conditions. Furthermore, interaction with real world and with human players is also seamlessly integrated in real-time, as video games are designed to facilitate the prompt interaction between human players and non-player characters (NPC).

While old game character implementations (for instance, Pac-Man ghosts or Space Invaders alien spacecrafts) were based on really simple pre-programmed and scripted behaviours, modern AAA video games are developed to simulate real complex environments and they require engaging, realistic and believable human-like behaviour for their NPCs. Although scripted behaviours might still be acceptable for some specific scenarios, AAA game consumers expect to find synthetic characters at the same level of behavioural realism and unpredictability as evoked by the visual experience of the game.

Generally, human-like behaviour is difficult to both define and test. In fact, the Turing test paradigm stills apply to this problem because no better alternatives have been found to characterise human behaviour. In the realm of computer games, this elusive characterisation might, in principle, be seen easier to define. For instance, human players usually consider disappointing the behaviour of artificial characters for two main reasons (Nareyek, 2004): they are either too intelligent, rational and accurate to be human, or on the contrary, they are too silly. Therefore, the challenge is to find that blurred medium level that characterises human player behaviour.

From the point of view of cognitive science, human-level intelligence and human-like behaviour can be considered as produced by several interrelated psychological processes, ranging from basic activation processes like primary motivations to complex high level cognitive processes such as set shifting and imitation learning. The current knowledge we have about these processes can be used to inspire the design of artificial cognitive architectures. In this paper, we present three different approaches to this sort of inspiration and put them to the test in an adapted version of the Turing test based in a video game (Hingston, 2009). Additionally, we assess the believability (or "humanness") of these bots using two different assessing methods: First-person and Third-person judges.

The remainder of this paper is structured as follows. In the next section we discuss the problems of assessing believability and describe the testing protocols we have used in this research. In Section 3 we present the different approaches to the design of believable agents, followed in Section 4 by a description of the implementations that we have developed for the believability experiments. Finally, experimental results are presented in Section 5 and discussed in Section 7.

2. Testing for believability in video games

Testing for human-like behaviour is not straightforward as different observers usually pay attention to different aspects (Arrabales et al., 2012). Therefore the task of judging the believability of a video game character can be approached from the perspective of inter-subjective assessment. In this context there is a key factor to take into account: the possible differences between First-person and Third-person observation. Togelius et al. (2012) argue that believability is better assessed from a Third-person perspective rather than a First-person perspective, i.e. where the assessor is not a participant in the game. As described below, the BotPrize testing protocol forces all human judges to take active part in the game and perform the assessing task as First-person observers. One of the main contributions of this work is to compare the believability results of the same bots both using the First-person perspective of the BotPrize environment and the Third-person perspective using recorded video from the very same testing sessions. In other words, believability for each bot is assessed using two different methods but using the same game play data.

2.1. The BotPrize testing protocol

The first method that we have used in order to assess the believability of our bots is the international BotPrize competition testing environment (Hingston, 2009). The BotPrize challenge (held yearly since 2008) was originally conceived as a Turing test for First-Person video game bots (NPCs). In the classical Turing test accurate verbal report and conversational skills are the key factors, however in the domain of FPS bots these aspects are neglected, focusing the assessment completely in observed non-verbal sensory-motor skills.

BotPrize environment is based in the video game "Unreal Tournament 2004" by Epic Games, a First-Person Shooter set in a fictional future with futuristic weapons. The objective of the game (deathmatch mode) is to kill as many opponents as possible without being killed by the other players. Both artificial bots and human players connect to the game server by means of a local area network or over the Internet.

Different judging schemes were used in early editions of the BotPrize competition. In this work, we use the latest scheme adopted in 2010 (Hingston et al., 2010). In this scheme a judging gun (the "Link Gun") is included in the game. All players, humans and NPCs spawn with a Link Gun with infinite ammo. Although the primary and alternate fire modes of the judging gun look and sound the same to all observers, they have completely different meanings and effects: the primary firing mode is meant to issue a vote for a bot (artificial player or NPC) and the alternate firing mode is meant to issue a vote for a human player.

If a human player shoots a bot or another human using the primary firing mode of the Link Gun, then the bot or the other human player obtains a bot vote. Analogously, If a human player shoots a bot or another human using the alternate firing mode of the Link Gun, the other player (bot or human) obtains a human vote. At the end the humanity percentage will be the human votes divided by the total number of votes received.

During our tests we allow the players to judge any other players as many times as desired. Using the judging gun the game play is transformed from a pure deathmatch game into a hybrid game in which both judging and killing/surviving aspects have to be taking into account simultaneously. It is important to remark that all players (humans and artificial) have access to the judging gun. Therefore, the designer of a bot also has to take decisions on how and when the bot will use the Link Gun, as this usage will also be part of the observed behaviour.

Apart from the judging gun, the rest of the weapons function as usual. However, the damaged produced by these weapons is reduced by a 60%, thus giving human players more chances to observe the other players before being riddled under enemy fire.

In order to obtain a significant amount of judging data and reduce the bias that a given map would introduce, different testing sessions in different maps are organised using a centralised game server that runs the BotPrize mod. Each session lasts for 15 min and different maps and scenarios are used each time. Anonymity of players is guaranteed using random player names and random player skins (clothes and body appearance) that changes from one session to the next.

The number of human players and bots is balanced, having a similar number of human judges and artificial characters. All human players are meant to be judges, but they also compete for the highest score (that they obtain both from judging and from killing and not being killed).