Design of a Teammate AI by Learning Human-player Utility from a few Records of Actions

Wada Takayuki(1310082)

School of Information Science,
Japan Advanced Institute of Science and Technology

February 12, 2015

Keywords: Game AI, RPG, Utility, Machine Learning, Team-mate, Cooperation Game.

One of the goals of game AI research is to create computer players (hereinafter referred to as game AIs) able to entertain human players. Until now, most researchers in that area tried to entertain human players by creating sufficiently strong AIs. As a result, game AIs are strong enough to compete with professional players in Othello, chess or shogi. Additionally, creating game Ais in many popular computer games is also a subject of study, like FPS games or RTS games. Methods to create game AIs that are strong or that behave naturally have been proposed. However, there are few researches for game AIs to entertain human players as their teammates. Some genres of commercial computer games, especially RPG games, often allow players to play the game with the game AIs as the teammates. To entertain human players in these games, it is important to make the action patterns of the teammate AIs more sophisticated than the opponent AIs. These teammate AIs are required to cooperate with human players but the AIs often take actions that human players do not expect them to do. Such mismatches between the expectations of the human players and the actions taken by the AI players often cause dissatisfaction of the players. One of the reasons for such mismatches is that there are several types of sub-goals in these games, like "win as soon as possible" or "win with as little received damage as possible", and the AI players act without understanding which types of sub-goals are important for each human player. The purpose of this study is to propose a method to develop teammate AI players that estimate the sub-goal preference of the human players and act with causing less dissatisfaction of the players. First of all, we designed and implemented a RPG game with rules suitable for this study. We picked out the design of role-playing video games famous recently in Japan, in which players select actions in

their turn, from among several types of RPG games. The situations in the game are unfair and mostly advantageous for human players. However, if the player wins, the last state of the game is sometimes preferable or unpreferable (e.g. some of the teammates are badly injured or no "Magic power" remains) for the players. These preferences for some game states are not expressed in numerical form. These game features are not rare in such games and are also included in our implemented game. In this study, we construct a model of such human preference with a function. The input of the model function is the game state and the selected actions, and the output is a number quantifying the preference. By adjusting the variable parameters in this model function, we mimic the preference for the game states of each human player to some extent. Additionally, we construct AI players that select actions using the adjusted model function. These AI players have the same preference for the game states as the human players and act with causing less dissatisfaction of the players. To estimate the human's preference, we use Monte-Carlo simulations and simulate how the game state will be at the end of the game by the human player's selected action. The performance of Monte-Carlo simulations often gets improved by selecting more frequently plausible actions. Although, the plausible actions by human players vary according to the sub-goals that the players have. Thus we defined a "strategy" (a strategy makes the action patterns of the player biased in the random simulations) and simulated the game states applying some strategies to the players in the random simulations. This approach succeeded in improving the accuracy of the estimation of the human player's preferences. We did an evaluation experiment with artificial human players that select their actions by fixed value functions and we tried to estimate the preferences of these artificial human players by our proposed method. The selected actions based on the estimated preferences were the same as the selected actions by the original artificial players at the rate of 67.1% in one setting. The upper bound of the rate is about 70.6%(in this setting), which is the rate at which the same actions are selected when the preferences are exactly the same. Thus the proposed method is only 3.5% inferior in performance in the worst case compared to an ideal estimation. Moreover, we did an evaluation experiment with human players and tried to estimate the preference of the players. At first, we gave them orders, like "fight trying to save Magic Powers of the characters" or "try to win the game as soon as possible" before the games. We estimated the preferences of the players through 4 games. Secondly, each human player played the game with a game AI as a teammate. There are two types of these teammate AIs, that is, AIs selecting their actions by the estimated value functions, or some fixed value functions that is unrelated to the human's preference. The human likeness of their behavior is ranked on a five grade scale. The proposed method got higher grade (0.7 on average), thus, our proposed method succeeded in estimating the human players' preference for the

game states and acted without giving the human players uncomfortable feelings.