Bomberman learning

A few weeks ago, we participated in a computer science course about Reinforcement Learning, as a final project in the course, some students designed applications or games to promote the application of knowledge of this topic. This is our case, in which we studied a common game, the Bomberman , actually a simplified version of it, where the agent (the bomberman) has to escape from his world while destroying walls finding the path to the exit. To do this, he can walk around the map using his bombs (and his sophisticated detonator) trying not to kill himself.

In reinforcement learning, is important to know that the agent must learn by himself what the way to the best benefits is and not supervising it like other machine learning techniques; this means that during the learning process, there is nobody telling the agent what actions should or shouldn’t be taken. He should explore uncharted territories and exploit the current knowledge.

In order to do this, we had to program a *model* for the environment representing the “real bomberman world”. There, the bomber can act performing actions and changing the state of the environment. Each state, represents a distinguish snapshot of the world.

To explore the world, one of the basic ways is having the agent performing actions in the real word and getting rewards (positive or negative) to get some information about the world where he moves, in this aspect, he can storage certain information about the model. A possible training consists in the agent constructing copies of parts of the real model of the world (Model Based methods) and another possibility is just preserving the information about states, action taken and obtained rewards (Model Free).

Then, after programming the algorithms, the bomberman was trained using both methods acting in the world several times and the results were the following:

<object width="480" height="385"><param name="movie" value="http://www.youtube.com/v/Sp1CmOaSugU?fs=1&amp;hl=es\_ES"></param><param name="allowFullScreen" value="true"></param><param name="allowscriptaccess" value="always"></param><embed src="http://www.youtube.com/v/Sp1CmOaSugU?fs=1&amp;hl=es\_ES" type="application/x-shockwave-flash" allowscriptaccess="always" allowfullscreen="true" width="480" height="385"></embed></object>

In this video, we can see four instances of training were the agent is learning different things. The first try is the worst one, the second and third ones were he is improving his actions and finally the forth one were the bomberman execute the optimal steps to get to the exit.

<https://rl-projects.googlecode.com/hg/> Mercurial repository containing reinforcement learning projects, developed within the scope of ECI 2010 (<http://dc.uba.ar/events/eci/2010/cursos/diuk>)

For more information about model based versus model free methods, see: <http://webdocs.cs.ualberta.ca/~sutton/papers/kuvayev-sutton.pdf>