Algorithm implementation

Goal of algorithm is to update the approximation of c(s, a), V(s~) and C(s) so that it become sufficiently good.

Add huge reward for invalid action.

Implementation detail:

* Use dictionary (similar to map in C++) for representing the estimations of the functions.

Some issues while implement:

* Determine a good value for delta in equation 17
* Determine a good learning rate in equation 18 (the two constraints for learning rate seem to contradict?)
* Value for alpha in equation 20.
* What is the terminal condition for the algorithm?
* Technical problem: Can’t register the ID for discrete environment.