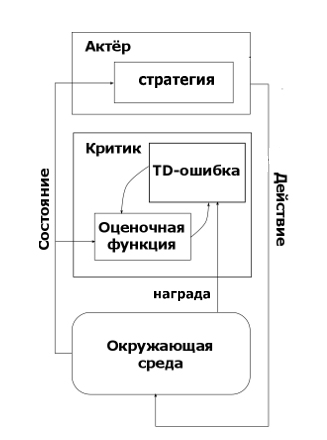
**2.4 Метод временной разницы (TD-метод)**

Большинство алгоритмов, названных в параграфе 2.1, опираются на так называемый TD-метод – метод временной разницы. [8] Он был предложен Саттоном и Барто в 1990 году в качестве альтернативы модели Рескорла-Вагнера, которая учитывает торможение и блокирование, а также предполагает, что обучение происходит только тогда, когда награда не равна ожиданию. [45] У этой модели отмечаются некоторые недостатки. Например, она не может описывать события второго порядка: если стимул А предсказывает стимул B, а стимул B предсказывает состояние S, то очевидна связь между A и S, что никак не отражается в этой модели. [29] TD-метод устраняет недостатки модели Рекорла-Вагнера. [29]

Целью TD-метода является оценка ситуации с точки зрения будущих наград. Происходит это следующим образом: состояние и подкрепительный сигнал (награда) передаются из окружающей среды ко критику. Критик оценивает состояние функцией и высчитывает ошибку предсказания , которая используется для обучения и у критика, и у актёра. К актёру передается значение . Актёр на основании полученных данных генерирует стратегию . [29] На рисунке 3 изображена схема TD-обучения.

Рисунок 3 – TD-метод



Рассмотрим этот процесс подробнее.

Различные состояния из множества состояний следуют друг за другом согласно некоторому заранее определенному распределению *.* Награда наблюдается в состоянии с вероятностью *.* Значением состояния называется математическое ожидание суммы всех будущих наград.

Ожидание здесь является отношением между вероятностью перехода от одного состояния к другому и вероятностью вознаграждения в каждом состоянии. Тогда:

Согласованность между последовательными состояниями лежит в основе TD-метода. Последовательность имеет место лишь для тех значений, которые правильно прогнозируют ожидаемое . Если значения не верны, и равенство не выполняется, то говорят об ошибке предсказания временной разницы (TD-ошибка).

где – скорость обучаемости, а – TD-ошибка и

Таким образом, получаем:

С помощью TD-метода можно прогнозировать события, даже если среда не известна. [29]