**2.3 Обзор методов решения задачи**

Саттон и Барто выделяют три фундаментальных метода решения задачи обучения с подкреплением: динамическое программирование, метод Монте-Карло и метод временной разницы (Temporal Difference, TD-метод). [35]

Методы динамического программирования предполагают, что модель окружающей среды идеальна и выполняется марковское свойство. Это предположение является следствием того, что применяются такие методы очень редко. Основная идея таких методов состоит в том, что функция ценности используется для создания и поиска оптимальных стратегий. [35] Алгоритм можно описать следующим образом:

1. Выбирается стратегия
2. До тех пор, пока стратегия не устойчива:
   1. Стратегия оценивается, – функция оценки стратегии
   2. Стратегия улучшается, – улучшенная стратегия,

Методы динамического программирования могут быть эффективными для определенных задач.

Методы Монте-Карло основываются на осредненной выгоде для некой выборки. Их используют для заданий, которые можно разбить на эпизоды. Здесь оценивание происходит не на каждом шаге, а после завершения очередного эпизода.

Методы Монте-Карло имеют преимущества перед методами динамического программирования. Например, они менее чувствительны к выполнению марковского свойства. [35]

Методы временной разницы в настоящее время являются наиболее важными среди всех методов обучения с подкреплением. [35] Они объединяют в себе идеи методов динамического программирования и Монте-Карло. Подробнее TD-метод рассмотрен в параграфе 2.4.