Билет 22

Динамическое программирование. Принцип оптимальности. Уравнение Беллмана

Определение. *Динамическое программирование(ДП)* – это метод нахождения оптимальных решений в задачах с многошаговой структурой.

Динамика задач ДП заключается в методе решения.

Определение. *Управление* – это воздействие, переводящее систему из начального состояния в конечное.

Особенности задач ДП

1. Рассматривается система, состояние которой на каждом шаге описывается n-мерным вектором

$$x_t = (x_1(\mathbf{t}),\ x_2(\mathbf{t}),\ \dots\ ,\ x_n(\mathbf{t})).$$

Предполагаем, что время t изменяется дискретно и принимает целочисленные значения 0, 1, 2, ... Дальнейшее изменение ее состояния зависит только от данного состояния x_t и не зависит от того, каким путем система пришла в него (процессы без последствия).

2. На каждом шаге выбирается одно решение u_t, под действием которого система переходит из предыдущего состояния x_{t-1} в новое x_t (u_t - m-мерный вектор управления). Здесь

$$x_t = x_t(x_{t-1},\ u_t).$$

3. Действие на каждом шаге связано с определенным выигрышем или потерей, которые зависят от состояния на начало шага и принятого решения.

4. На векторы состояния и управления могут быть наложены ограничения, объединение которых образует область допустимых решений $\mathbf{u} \in \Omega$.

5. Требуется найти такое допустимое управление u_t для каждого шага t, чтобы получить максимальное значение функции цели за все T шагов.

Определение. *Стратегия управления* – это любая допустимая последовательность действий для каждого шага, переводящая систему из начального состояния в конечное.

Определение. *Оптимальная стратегия управления* – это допустимая стратегия управления, доставляющая функции цели экстремальное значение.

Принципы оптимальности

Любую многошаговую задачу можно решить различными методами. Например, можно искать сразу все элементы решения на всех N шагах. Этот метод не всегда приводит к цели, особенно тогда, когда целевая функция задана в виде таблиц или число переменных очень велико. Второй путь основан на идее проведения оптимизации поэтапно. Но оптимизация не предполагает изолированности, наоборот, управление на каждом шаге выбирается с учетом всех его последствий. Идея пошаговой оптимизации составляет суть метода ДП.

Из качественного анализа идеи поэтапной оптимизации можно сформулировать следующие принципы, лежащие в основе ДП.

Принцип оптимальности. Каковы бы ни были начальное состояние и начальные решения, последующие решения должны приниматься исходя из оптимальной стратегии с учетом состояния, вытекающего из первого решения.

Принцип погружения. Форма задачи, допускающей использование метода ДП, инвариантна относительно N.

Реализация данных принципов дает гарантию того, что решение, принимаемое в очередном шаге, окажется наилучшим относительно всего процесса в целом, а не узких интересов данного этапа.

Функциональные уравнения Беллмана

Так как оптимальная стратегия одновременно оптимальна для любого количества оставшихся шагов, ее можно строить по частям: сначала для последнего этапа, затем для двух последних и т.д., пока не придем к первому шагу.

Определение. *Условно-оптимальное управление (решение)* – это набор оптимальных управлений, зависящих от возможных исходов предыдущего этапа.

Математическая формулировка принципа оптимальности для задач с аддитивным критерием оптимальности

Будем считать, что начальное состояние x_0 и конечное состояние системы x_T заданы.

Обозначим $z_i(x_{i-1},\, u_i)$, $i=\overline{1,N}$ - соответствующее значение функции цели на i-ом этапе. Тогда

$$Z{=}z(x_0,\, u){=}\sum_{i=1}^{n} z_i(x_{i-1},\, u_i).$$

Надо найти оптимальное управление $u^*{=}(u_1^*,\, u_2^*,\, ...\, u_N^*)$, которое доставляет экстремум целевой функции при ограничениях $u \in \Omega$.

Для решения данной задачи мы погружаем ее в семейство подобных. Пусть Ω_i - область определения задачи на i-ом этапе. Обозначим через

$$F_1(x_{N-1}),\ F_2(x_{N-2}),\ ...\ ,\ F_N(x_0)$$

условно-оптимальные значения функции цели на последнем этапе, на двух последних, и т.д, на k последних и т.д, и на всех N этапах.

Начинаем с последнего этапа. Пусть x_{N-1}- возможные начальные состояния системы на начало N-го этапа. Находим

$$F_1(x_{N-1}){=}\underset{u_N\in\Omega_N}{max(min)} z_N(x_{N-1}, u_N).$$

Для двух последних этапов получаем

$$F_2(x_{N-2}){=}\underset{u_{N-1}\in\Omega_{N-1,N}}{max(min)}\,(Z_{N-1}(x_{N-2},\, u_{N-1}){+}F_1(x_{N-1})).$$

Аналогично:

$$F_3(x_{N-3}){=}\underset{u_{N-2}\in\Omega_{N-2,N-1,N}}{max(min)}\quad(Z_{N-2}(x_{N-3},\, u_{N-2}){+}F_2(x_{N-2})),$$

....

$$F_N(x_0){=}\underset{u_1\in\Omega}{max(min)}(Z_1(x_0,\, u_1){+}F_{N-1}(x_1)).$$

Данные выражения называются *функциональными уравнениями Беллмана.* Отчетливо просматривается их реккурентный характер, поэтому данные уравнения называются еще и реккурентными.

Проделав такой поиск условно-оптимальных решений для каждого шага от конца к началу, найдем последовательность условно-оптимальных управлениц $u_1^*(x_0)$, $u_2^*(x_1)$, ... , $u_N^*(x_{N-1})$. Найденное при этом $u_1^*(x_0)$ является оптимальным в силу того, что начальное состояние x_0 определятся однозначно. Зная его, выявим, к какому состоянию перейдет система в результате этого действия, т.е. находим оптимальное состояние системы x_1^* на начало второго этапа. Зная его, можно найти управление $u_2^*(x_1)$ для второго этапа (так как для всех возможных состояний второго этапа уже выявлены оптимальные управления). Таким образом, проделав обратное движение по условно-оптимальным управлениям от начала к концу, найдем просто оптимальное управление для всех этапов.