Задача оптимального управления в системах искусственного интеллекта с обратной связью

Андрей Сергеевич Веприков
Научный руководитель: д.ф.-м.н. А. С. Хританков

Кафедра интеллектуальных систем ФПМИ МФТИ
Специализация: Интеллектуальный анализ данных

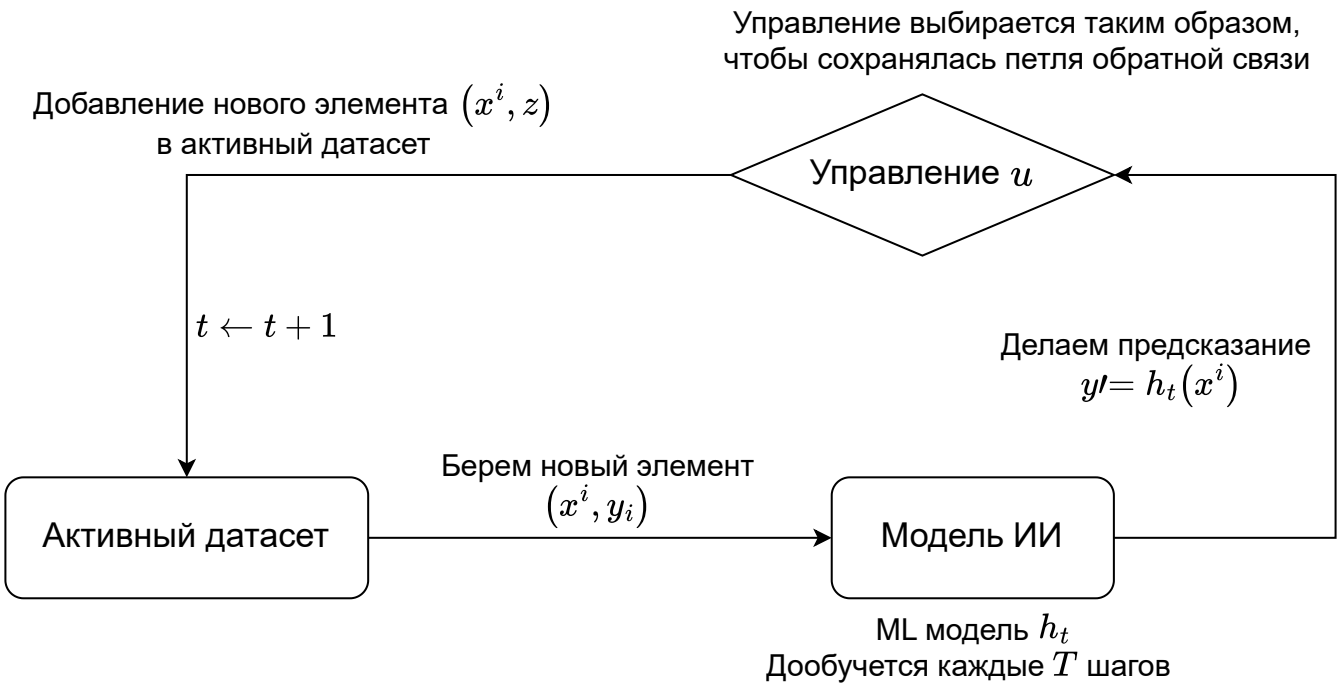
Декабрь 2024

Примеры эффектов петель обратной связи

1. Самоисполняющееся пророчество (self-fulfilling prophecy)
2. Вынужденное смещение данных (data drift) в рекомендательных системах
3. Усиление ошибок (error amplification) со временем в задаче медицинского прогнозирования
4. Дрейф данных в системах предиктивного полицейского контроля
5. Пузыри фильтров (filter bubbles)

Задача оптимального управления в системах искусственного интеллекта с обратной связью

Пусть дано множество функций плотности состояний системы $\mathcal{F}$, в которое нам необходимо попасть при $t \to \infty$. На каждом шаге t доступна ограниченная выборка данных системы и управление $u_t \in \mathcal{U}$.



Модельная задача оптимального управления в системе ИИ с петлей обратной связи.

Постановка задачи в терминах (Po)MDP

Определим пятерку: $(\mathcal{S}, \mathcal{U}, \mathbf{D}, \rho, \gamma)$, где

1. $\mathcal{S}$ – состояния
2. $\mathcal{U}$ – действия
3. $\mathbf{D}$ – оператор эволюции
4. r – награда
5. $0 < \gamma < 1$ – дисконтирующий коэффициент

Задача ставится как

$$\max_{u_1,\ldots,u_\infty} \sum_{t=0}^{+\infty} \gamma^t r\left(s_t, u_t\right), \tag{1}$$

при условии

$$s_{t+1} = \mathbf{D}(s_t, u_t). \tag{2}$$

В нашей задаче вводится ряд уточнений по сравнению с постановкой MDP:

1. **Ограничение на управление на каждом шаге**
2. **Оператор эволюции разбивается на 2 части: ответ пользователей и обучение агента**

Математическая постановка задачи

1. $(f_t, h_t) \in \mathcal{S}$ – функции плотности состояний системы и распределения весов ML модели
2. $u_t \in \mathcal{U}$ – доступные управления

3(a). $\mathrm{P}_t \in \mathbf{P}$ – оператор эволюции плотностей состояний системы f_t (решения пользователей)

3(b). $\mathrm{H}_t \in \mathbf{H}$ – оператор эволюции плотности весов ML модели h_t (алгоритм обучения)

4. $\rho(f_t, \mathcal{F})$ – функция расстояния до желаемого множества $\mathcal{F}$ (отрицательная награда)
5. $0 < \gamma < 1$ – дисконтирующий коэффициент
6. $g(f_t, h_{t+1})$ – функция, показывающая, что в системе с данными $\sim f_t$ и весами модели $\sim h_{t+1}$ сохранится петля обратной связи

Что нам не доступно	Что нам доступно
$f_t \in \mathcal{S}$	Ограниченная выборка из $f_t \in \mathcal{S}$
$\rho\,(f_t, \mathcal{F})$	Оценка расстояния $\rho(\tilde{f}_t, \mathcal{F})$
$\mathrm{P}_t \in \mathbf{P}$	$\mathrm{H}_t \in \mathbf{H}$, $u_t \in \mathcal{U}$, g, γ

Математическая постановка задачи

Оптимизационная задача:

$$\min_{u_1,\dots,u_\infty} \sum_{t=0}^{+\infty} \gamma^t \rho\left(f_t, \mathcal{F}\right), \tag{3}$$

при условии

$$h_{t+1} = \mathsf{H}_t\left(f_t, h_t, u_t\right), \quad \mathsf{H}_t \in \mathbf{H}, \tag{4}$$

$$f_{t+1} = \mathsf{P}_t\left(f_t, h_{t+1}\right), \quad \mathsf{P}_t \in \mathbf{P}, \tag{5}$$

$$g(f_t, h_{t+1}) \leq 0, \tag{6}$$

Уравнение (6) может быть переписано в виде $\mathbb{P}\{g(\mathbf{x}, \boldsymbol{\theta}) \leq 0\} \geq 1 - \delta$, где $\mathbf{x} \sim f_t$ и $\boldsymbol{\theta} \sim h_{t+1}$, то есть:

$$\int_{\mathbf{x}} \int_{\boldsymbol{\theta}} \mathbf{1}\{g(\mathbf{x}, \boldsymbol{\theta}) \leq 0\} f_t(\mathbf{x}) h_{t+1}(\boldsymbol{\theta}) d\mathbf{x} d\boldsymbol{\theta} \geq 1 - \delta. \tag{7}$$

Получили задачу стох. управления с интегральными ограничениями

Специфика интеллектуальной системы с петлей обратной связи

Жадный алгоритм управления через ω-предельное множество

1. Пусть мы находимся на шаге t. Нам доступны $\hat{f}_t$, h_t и $\mathbf{H}$.
2. Из всех возможных управлений $u_t \in \mathcal{U}$ выбираем те, которые сохранят петлю обратной связи, то выполнено ограничение (7)
3. Для каждого подходящего управления находим, какие распределения весов $\{h^{\infty}_{u_t}\}$ могут быть у модели в пределе, если на шагах $t+1, ..., \infty$ мы не будем управлять.
4. Для задачи обучения с учителем, так как в системе присутствует петля обратной связи, плотность данных $f^{\infty}_{u_t}$ для управления u_t в пределе будет иметь вид $f^{\infty}_{u_t}(\mathbf{x}, \mathbf{y}) > 0 \Leftrightarrow y = h^{\infty}_{u_t}(\mathbf{x})$ для всех примеров состояния системы $(\mathbf{x}, \mathbf{y})$
5. Выбираем u_t, которое максимизирует меру пересечения между $\{f^{\infty}_{u_t}\}$ и $\mathcal{F}$
6. Переходим на шаг $t+1$

Специфика интеллектуальной системы с петлей обратной связи

Как мы можем управлять агентом?

1. Замена некоторых элементов тренировочной выборки (или только целевой переменной y)
2. Замена модели ML
3. Adversarial атаки

Предлагаемые способы решения

1. PoMDP [Kaelbling et al., 1998]
2. Гауссовские фильтры (Фильтр Калмана) [Patil et al., 2015]
3. Задача оптимального управления в дискретных системах с интегральными ограничениями [Kamien, 2012]
4. Динамическое программирование [Bertsekas, 2012]
5. Стабилизирующее управление [Lechner et al., 2022]

Список литературы


Bertsekas, D. (2012).

Dynamic programming and optimal control: Volume I.


Kaelbling, L. P., Littman, M. L., and Cassandra, A. R. (1998).

Planning and acting in partially observable stochastic domains.


Kamien, M. (2012).

Dynamic Optimization: The Calculus of Variations and Optimal Control in Economics and Management.


Lechner, M., Žikelić, Đ., Chatterjee, K., and Henzinger, T. A. (2022).

Stability verification in stochastic control systems via neural network supermartingales.


Patil, S., Kahn, G., Laskey, M., Schulman, J., Goldberg, K., and Abbeel, P. (2015).

Scaling up gaussian belief space planning through covariance-free trajectory optimization and automatic differentiation.