

# ДИНАМИЧЕСКАЯ СЕГМЕНТАЦИЯ ПРОСТРАНСТВА ПРИЗНАКОВ ДЛЯ СИСТЕМ АВТОНОМНОГО АДАПТИВНОГО УПРАВЛЕНИЯ И СИСТЕМ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ

*А. Е. Лебедев, А. А. Жданов*

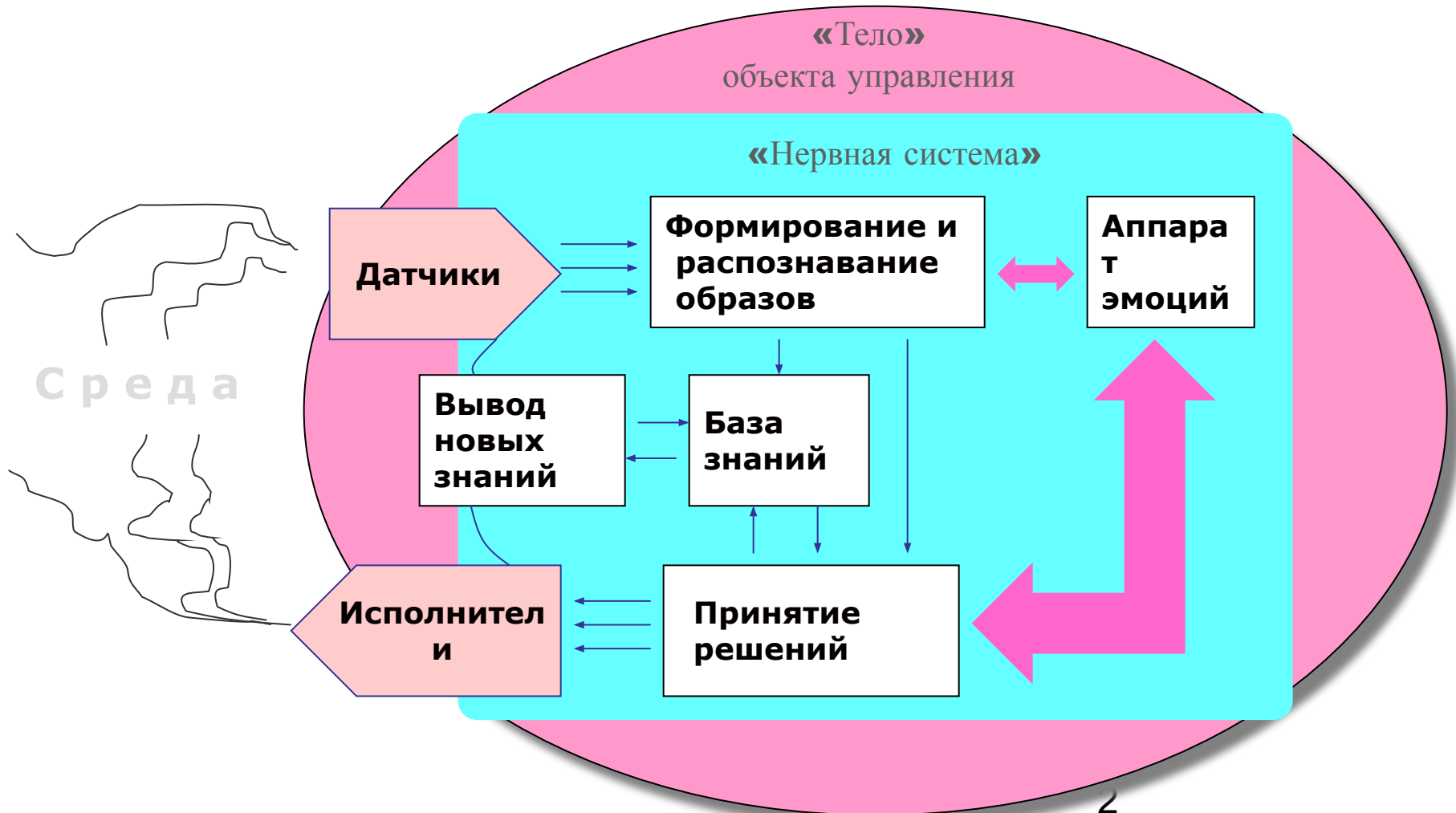
Институт точной механики и вычислительной техники имени  
С.А. Лебедева, Москва

Autonomous Adaptive Control Lab (AAC Lab)

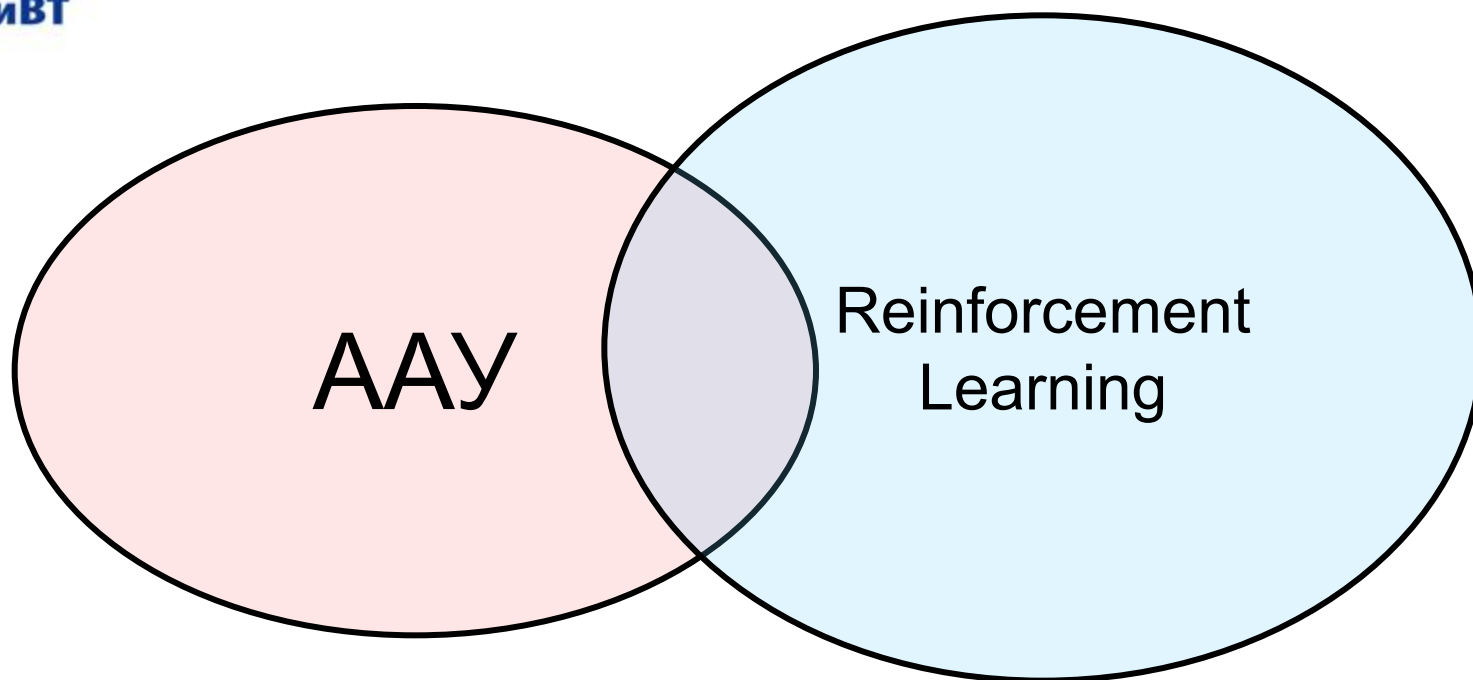
<http://www.ipmce.ru>  
<http://www.aac-lab.com>

# Структура и функции «нервной системы» Автономного Адаптивного управления

(ААУ)  
(показаны только основные связи)



# Соотношение методов ААУ и обучения с подкреплением



Образы - условия  $\longleftrightarrow$  Состояния

Эмоциональная оценка  
результата действия  $\longleftrightarrow$  Подкрепление  
(вознаграждение)

# Аппроксимация vs. Дискретизация

Набор входных параметров:  $p_1 \dots p_n$

Множество состояний  $\{s_i\}$

Набор доступных действий  $a_1 \dots a_k$

Оценка ожидаемого подкрепления  $Q(s_i, a_i)$

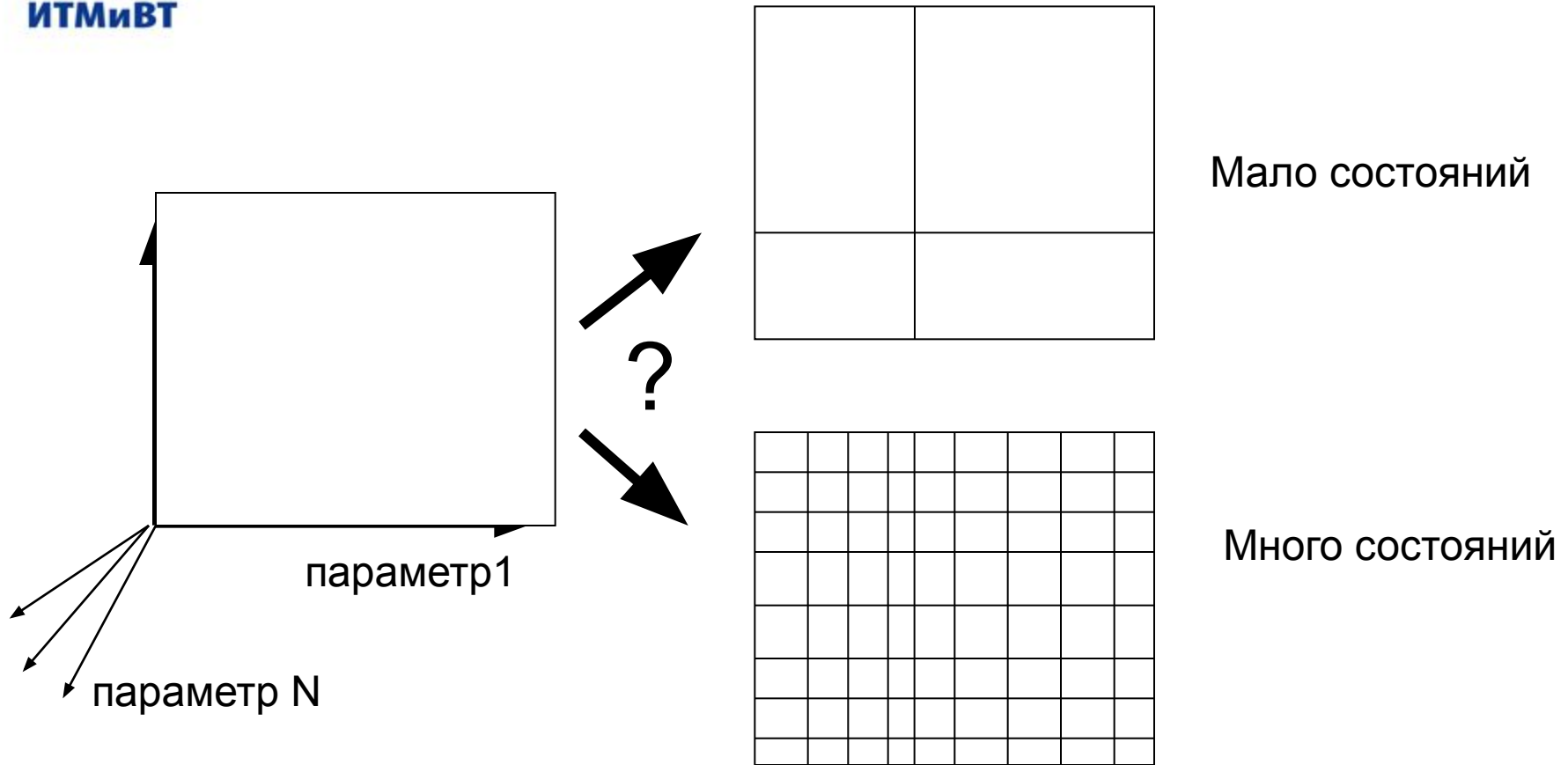
- Аппроксимация

$$\text{Est}(p_1 \dots p_n, a_i): \{(p_1 \dots p_n)\} \times a_i \rightarrow \mathbb{R}$$

- Дискретизация

$$\{(p_1 \dots p_n)\} \rightarrow \{s_i\},$$

$$\text{Est}(s, a): \{s_i\} \times \{a_i\} \rightarrow \mathbb{R}$$



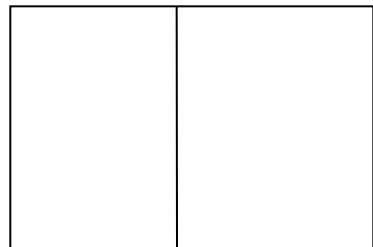
- При малом числе состояний низкая точность, при большом – долгое обучение.
- Кроме того, при любом фиксированном числе градаций для каждого из параметров число состояний растет экспоненциально при линейном росте числа параметров. («комбинаторный взрыв»)

# Предлагаемое решение

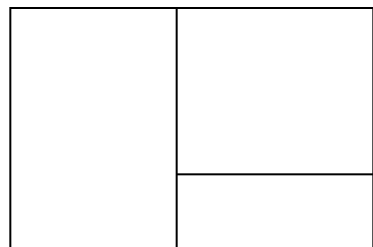
последовательная сегментация пространства признаков в процессе обучения систем



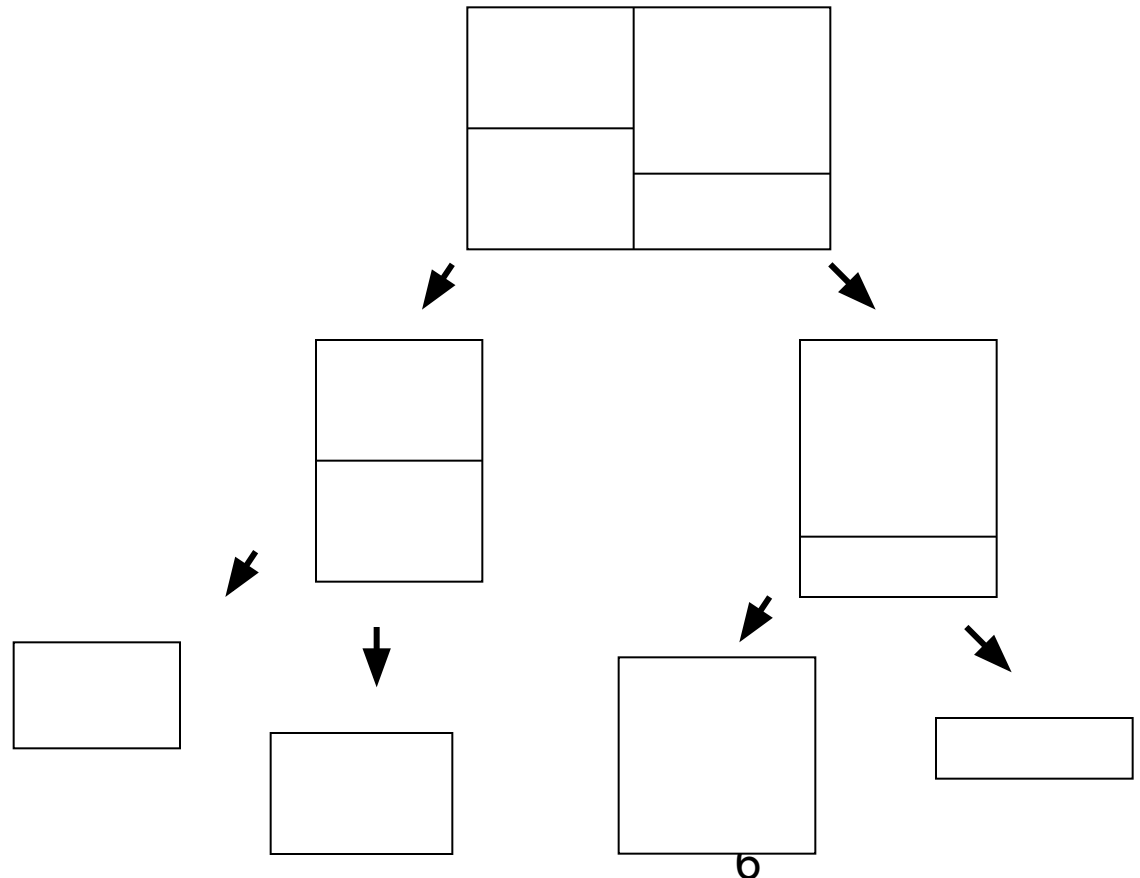
↓ обучение



↓ обучение



В результате образуется дерево вложенных состояний





ИТМИПТ

# Основные подзадачи



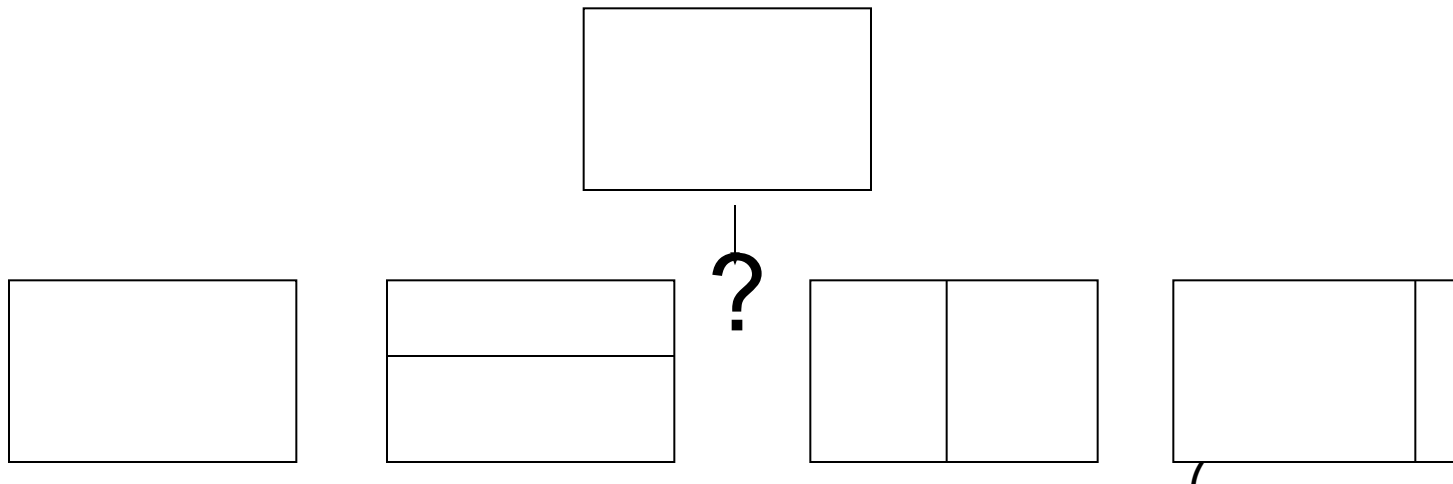
- Производить ли разделение очередного состояния на более мелкие и когда?

$$\text{Dev}(\text{Est}(s_{\text{old}}, a_{b0}) \geq \min(\text{Dev}(\text{Est}(s_{\text{new1}}, a_{b1}), \text{Dev}(\text{Est}(s_{\text{new2}}, a_{b2})))$$

$S_{\text{old}}$  — исходное состояние,  $S_{\text{new1}}$ ,  $S_{\text{new2}}$  — дочерние состояния  
 $a_{b0}$ ,  $a_{b1}$ ,  $a_{b2}$  — лучшие действия для соответствующих состояний

- Если да, то какой вариант разбиения выбрать? (как определить направление и пропорции?)

$$\max(\text{Est}(s_{\text{new1}}, a_{b1}), \text{Est}(s_{\text{new2}}, a_{b2})) \rightarrow \max$$



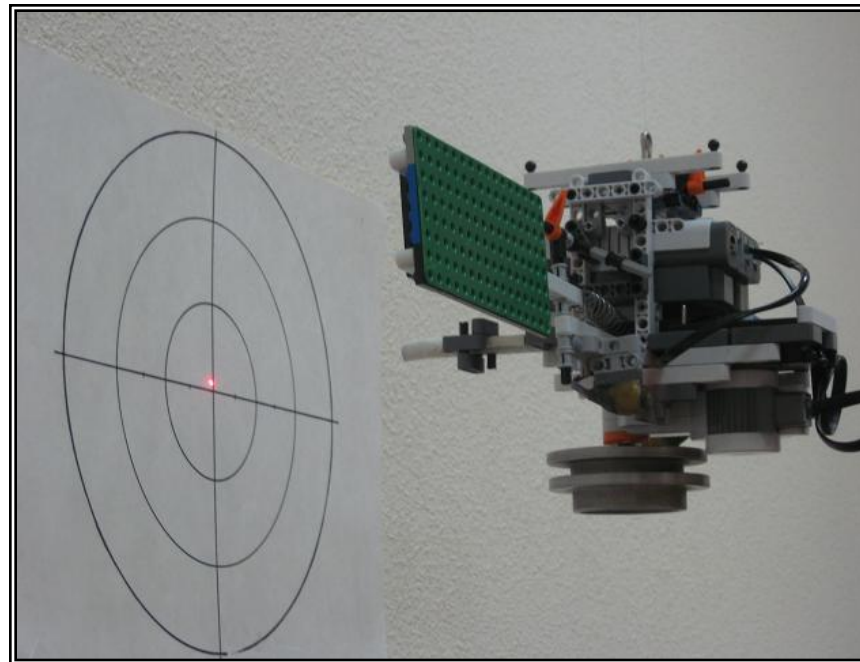
## Свойства разработанного метода

- Итоговое разбиение пространства признаков на состояния неравномерно и отвечает специфике конкретной задачи.
- Определенное улучшение качества работы возможно в самом начале обучения без накопления большой статистики
- Накопление статистики для различных состояний происходит независимо (отсутствует «катастрофическое забывание»)
- Получившийся в результате обучения закон управления может быть представлен в виде дерева решений и проанализирован человеком

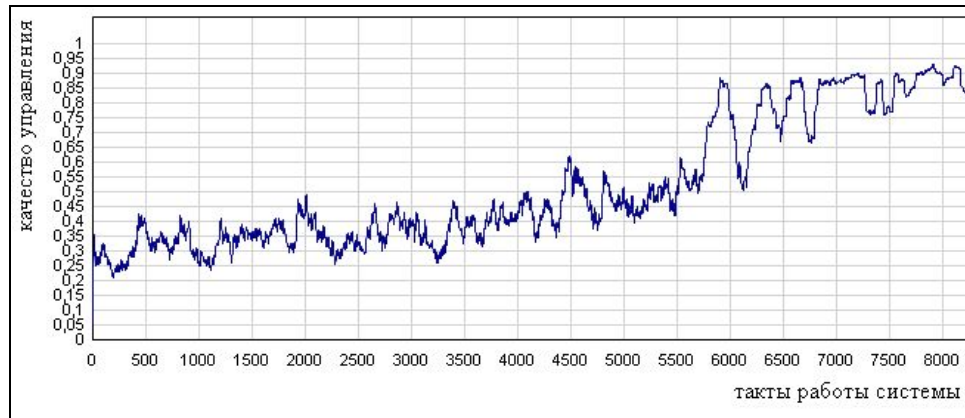


## Экспериментальное исследование

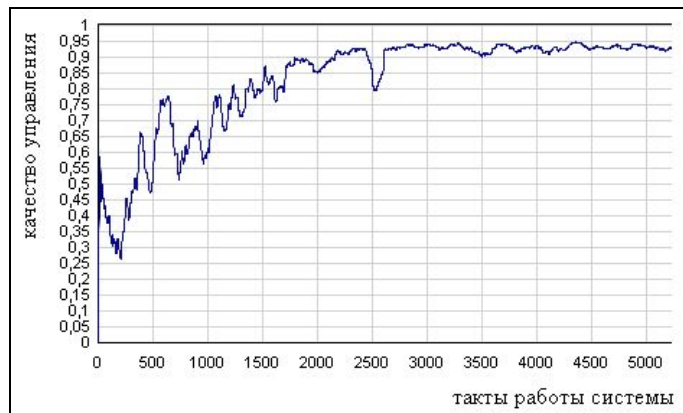
Экспериментальное исследование разработанного метода проводилось на компьютерной и на физической модели наноспутника с адаптивной системой управления



- Пример графика изменения качества управления при обучении с фиксированным набором состояний...



- И с использованием алгоритма динамической сегментации состояний



# Спасибо за внимание!

*А. Е. Лебедев, А. А. Жданов*

ИТМиВТ им. С.А. Лебедева, Москва  
AAC Lab

<http://www.ipmce.ru>  
<http://www.aac-lab.com>

[aazhdanov@ipmce.ru](mailto:aazhdanov@ipmce.ru)