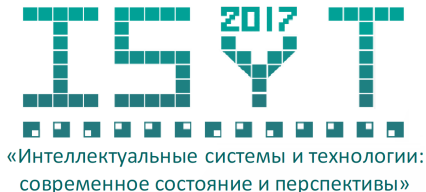




ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ  
БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ НАУКИ  
**ИНСТИТУТ  
ПРОБЛЕМ  
УПРАВЛЕНИЯ**  
ИМ. В.А. ТРАПЕЗНИКОВА  
РОССИЙСКОЙ АКАДЕМИИ НАУК



## Выявление предпочтений пользователя умного дома

Н.И. Базенков, М.В. Губко

[n.bazenzkov@yandex.ru](mailto:n.bazenzkov@yandex.ru)

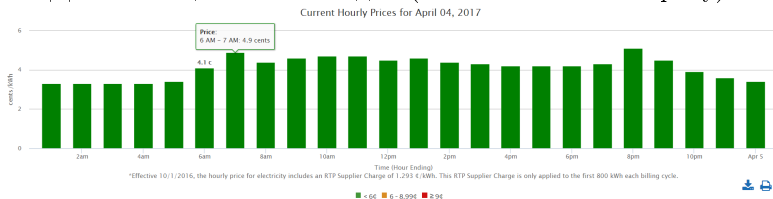
Институт проблем управления им. В.А. Трапезникова

2 июля 2017

# Введение

- 1 В электроэнергетике есть тенденция к либерализации цен для конечных потребителей
- 2 При динамическом ценообразовании планирование энергопотребления желательно автоматизировать
- 3 Автоматизация должна учитывать предпочтения пользователя
- 4 Появляются умные устройства, подстраивающиеся под привычки пользователя (термостат NEST).

## Динамика цен в течение дня (Ameren Illinois Company):



# Цель исследования

- 1 “Умный” электроприбор (пример - хлебопечка).
- 2 Когда запас хлеба подходит к концу, пользователь загружает необходимые ингредиенты и задает желаемое время завершения программы (обычно утро следующего дня)
- 3 Максимальный комфорт - завершить точно к завтраку
- 4 Минимальная цена - испечь ночью. Но к завтраку хлеб будет не такой вкусный



Прибор должен рекомендовать наилучшее время завершения с учетом текущей ситуации и предпочтений пользователя

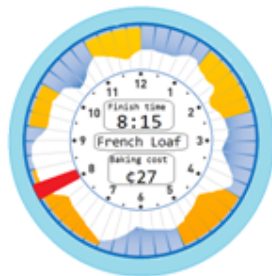
# План исследования

- 1 Имитационная модель **пользователя и прибора**.  
Пользователь моделируется как рациональный агент, учитывающий и комфорт, и цену
- 2 Два алгоритма **выявления предпочтений пользователя**:
  - 1 Параметрическая модель + байесовское обучение
  - 2 Обратное обучение с подкреплением (непараметрический метод)
- 3 **Эксперименты**:
  - 1 Цены по данным Pacific Gas and Electric Company
  - 2 Обучающие данные сгенерированы имитационной моделью.
  - 3 Сравнение с распространенными методами машинного обучения: случайный лес, kNN, SVM, XGBoost и др.

# Модель пользователя и прибора

- 1 Каждый день пользователь ест хлеб на завтрак, обед (если дома) и ужин.
- 2 Свежесть хлеба убывает со временем
- 3 Выбирая время, пользователь видит прогноз стоимости для каждого сценария
- 4 Комфорт – агрегированное “удовольствие” от всех приемов пищи

Прототип интерфейса:

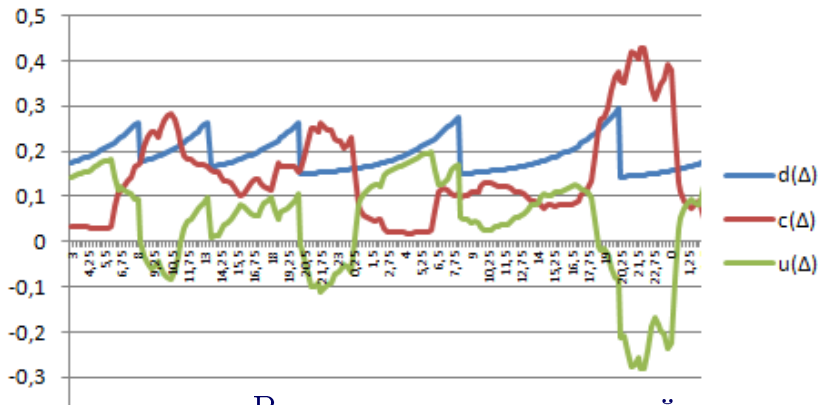


# Принятие решения пользователем

- 1 Выбор времени завершения программы  $\Delta$  в ситуации  $y$  максимизирует функцию полезности:

$$u(\Delta, y) = d(\Delta, y) - c(\Delta, y). \quad (1)$$

- 2  $d(\Delta, y)$  – уровень комфорта,  $c(\Delta, y)$  – стоимость для времени завершения  $\Delta$



# Задача выявления предпочтений

- 1 Известна история решений пользователя  $\{(y^i, \Delta^i)\}, i = \overline{1, n}$   
 Описание ситуации  $y^i$  включает:
  - Запас хлеба в момент выбора  $x_0^i$
  - День недели на момент завершения программы  $w^i$ : будний/выходной
  - Стоимости всех альтернатив  $c_1^i, \dots, c_N^i$  – каждый день поделен на  $N = 96$  интервалов по 15 минут.
- 2 Необходимо восстановить неизвестную функцию комфорта  $d(\Delta, y)$

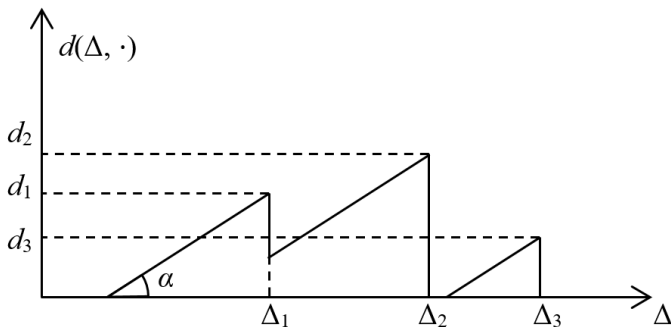
## Регрессионные методы VS Выявление предпочтений

- |   |  |
|---|--|
| <ol style="list-style-type: none"> <li>1 Черный ящик – нет внутренней модели пользователя</li> <li>2 Предсказывает действие напрямую на основе истории</li> </ol> | <ol style="list-style-type: none"> <li>1 Использует модель принятия решений</li> <li>2 Предсказывает будущие действия на основе модели</li> <li>3 Должно лучше обобщаться на данные, не включенные в обучающую выборку?</li> </ol> |
|---|--|

# Параметрическое байесовское обучение

- 1 Параметрическая модель функции комфорта
- 2 Генерация множества гипотез
- 3 Оценка правдоподобия каждой гипотезы для обучающей выборки, получение апостериорных вероятностей
- 4 Предсказание – взвешенный медианный сценарий

Модель функции комфорта:





# Генерация гипотез

- 1 Множество гипотез:  $(y, \omega) \in \Omega$ ,  $\omega = (d_1, d_2, d_3, \Delta_1, \Delta_2, \Delta_3, \alpha)$ .
- 2 Все возможные ситуации разбиты на 8 “базовых”  $y = (x_0, w)$ :
- 3 Запас хлеба  $x_0$  – 4 интервала:  $[0, 0.15)$ ,  $[0.15, 0.30)$ ,  $[0.30, 0.45)$ ,  $[0.45, 0.6]$
- 4 День недели  $w \in 0, 1$ ,  $1 =$  выходной,  $0 =$  будний

Запас хлеба	День недели	Параметры функции комфорта
$[0, 0.15)$	будний	$\omega^1 = (d_1, d_2, d_3, \Delta_1, \Delta_2, \Delta_3, \alpha)$
$[0, 0.15]$	выходной	$\omega^2 = (d_1, d_2, d_3, \Delta_1, \Delta_2, \Delta_3, \alpha)$
$[0.15, 0.30)$	будний	$\omega^3 = (d_1, d_2, d_3, \Delta_1, \Delta_2, \Delta_3, \alpha)$
	...	

# Выбор гипотезы

- 1 Дана гипотеза  $(y, \omega)$
- 2 Вероятность  $i$ -го наблюдения  $r^i = (y^i, \Delta^i)$  :

$$L(y, \omega | r^i) = \exp(-\gamma \pi(y, \omega | r^i)), \quad (2)$$

$\pi(y, \omega | r^i)$  – штраф за несовпадение с реальным выбором  $\Delta^i$

$$\pi(y, \omega | r^i) = K(y, y^i) \sum_{\Delta \in A} \max[0; u(\Delta, (y, \omega)) - u(\Delta^i, (y, \omega))] \quad (3)$$

- 3  $K_i = \exp(-\beta \rho(y^i, y))$  - ядерная функция.  
 Расстояние от ситуации  $y^i$  до “базовой”  $y = (x_0, w)$ :  
 $\rho(y^i, y^0) = \gamma |w^i - w| + |x_0^i - x_0|$ .  
 $\beta$  – чувствительность к расстоянию,  $\gamma$  – вес дня недели

# Предсказание действия

- ① Правдоподобие гипотезы  $(y, \omega)$  для всех наблюдений  $r^1, \dots, r^n$ :

$$L(y, \omega | r^1, \dots, r^n) = L(y, \omega | r^1) \times \dots \times L(y, \omega | r^n) \times L_0(y, \omega), \quad (4)$$

$L_0(y, \omega) = \exp(K_0 \pi_0(y, \omega))$  – априорное распределение

- ② Апостериорные вероятности  $p(\omega | y)$  – нормализация правдоподобия
- ③ Предсказание выбранного времени завершения – медианный сценарий:

$$\Delta^*(y^i, c^i(\cdot)) := \min \left\{ \Delta : \sum_{\omega: \Delta^*(y^i, c^i(\cdot), \omega) \leq \Delta} p(\omega | y^i) \geq 0.5 \right\} \quad (5)$$

- ④ Сценарий (5) минимизирует среднюю абсолютную ошибку (MAE):

$$\sum_{i=1}^n |\Delta^i - \Delta^*(y^i, c^i(\cdot))| \quad (6)$$

# Обратное обучение с подкреплением

Обучение с подкреплением (Reinforcement Learning) – играем на стороне агента

- 1 агент наблюдает реакцию среды на свои действия
- 2 надо найти оптимальную стратегию поведения

Обратное обучение с подкреплением (Inverse Reinforcement Learning) – играем на стороне “наблюдателя”

- 1 наблюдаем поведение агента
- 2 функция выигрыша неизвестна
- 3 надо найти функцию выигрыша, считая, что агент действует оптимально

# Алгоритм IRL

Критерий контраста [Ng, Russel 2000]:

$$\max_{y \in \mathcal{S}} [u(\Delta(y), y) - \max_{\Delta \in A \setminus \Delta(y)} u(\Delta, y)] \quad (7)$$

$$u(\Delta(y), y) \geq u(\Delta, y) \quad \forall \Delta \neq \Delta(y) \quad (8)$$

- $y$  – ситуация: запас хлеба и день недели
- $\Delta(y)$  – выбранное время завершения программы в ситуации  $y$
- $u(\Delta, y)$  – полезность действия  $\Delta$  в ситуации  $y$ :  
 $u(\Delta, y) = d(\Delta, y) - c(\Delta, y)$

# Алгоритм IRL

- 1 Все возможные ситуации разбиты на 8 “базовых”  $y = (x_0, w)$ :
- 2 Запас хлеба  $x_0$  – 4 интервала:  $[0, 0.15)$ ,  $[0.15, 0.30)$ ,  $[0.30, 0.45)$ ,  $[0.45, 0.6]$
- 3 День недели  $w \in 0, 1$ , 1 = выходной, 0 = будний
- 4 День разбит на  $N = 96$  интервалов по 15 минут.
- 5 Каждому интервалу в каждой “базовой” ситуации соответствует свое значение комфорта  $d(\Delta, y)$ :  
 $96 \times 2 \times 4 = 768$  переменных  $d_1(y), \dots, d_N(y)$ .
- 6 Для каждой базовой ситуации ищется своя функция комфорта как набор чисел  $d_1, \dots, d_N$

# IRL algorithm

Для каждой базовой ситуации  $y'$  ищем функцию комфорта как решение задачи:

$$\max_{d_1, \dots, d_N} \sum_{i=1}^n [d(\Delta^i) - \max_{\Delta \neq \Delta^i} \{d(\Delta) - c^i(\Delta)\}] K(y, y') - \alpha \|Ld\|^2 \quad (9)$$

$$d(\Delta) - c^i(\Delta) \leq d(\Delta^i) - c^i(\Delta^i) - \varepsilon \quad \forall \Delta \neq \Delta^i \quad (10)$$

$$d_k \geq 0 \quad (11)$$

- $d(\Delta)$  – комфорт при выборе действия  $\Delta$
- $\Delta^i$  – действие, выбранное пользователем в ситуации  $y^i$
- $d_k$  – оценка комфорта  $k$ -го дневного интервала в ситуации  $y$
- $c^i(\Delta)$  – затраты при выборе действия  $\Delta$
- $K_i = \exp(-\beta \rho(y^i, y'))$  – близость ситуации  $y^i$  к базовой  $y'$
- $\alpha \|Ld\|^2$  – регуляризация с матрицей  $L$ ,  $\alpha \in [0, 1]$ .

# Генерация данных

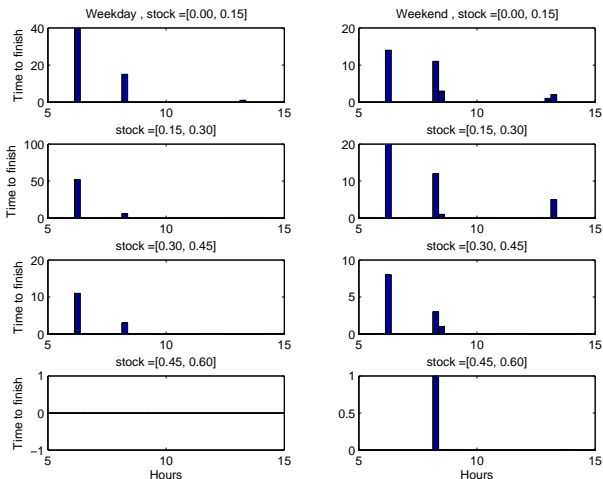
- 1 Обучающие и тестовые данные сгенерированы имитационной моделью пользователя
- 2 Наборы с разной дисперсией цен: низкая ( $L$ ), средняя ( $M$ ) и высокая ( $H$ ) price variance
- 3 Набор со средней дисперсией разбит на обучающую  $M^{80\%}$  и тестовую  $M^{20\%}$  выборки
- 4 Поиск оптимальных гиперпараметров  $\alpha, \beta, \gamma$  – кросс-валидация на  $M^{80\%}$

Дисперсия цен	Кол-во наблюдений	Мин. время завершения	Макс. время завершения
Низкая	163	6 <sup>00</sup>	13 <sup>00</sup>
Средняя	231	6 <sup>00</sup>	13 <sup>00</sup>
Высокая	127	6 <sup>00</sup>	13 <sup>15</sup>

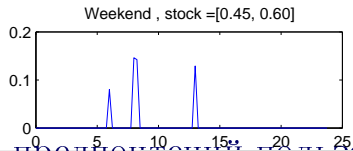
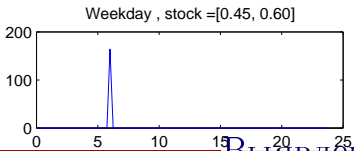
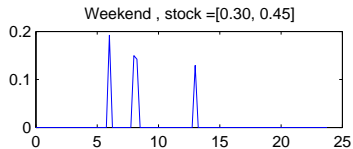
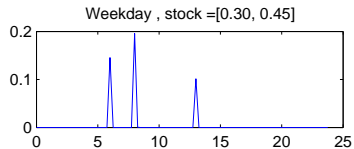
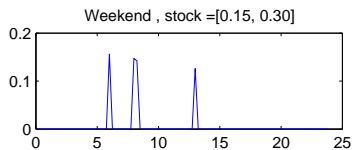
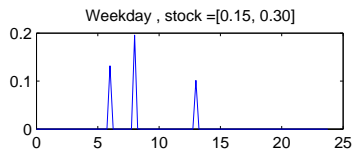
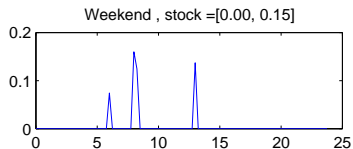
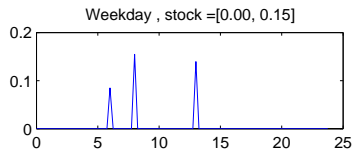


# Распределение времени завершения

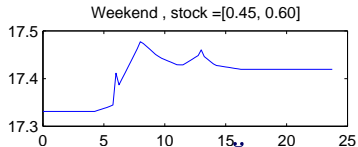
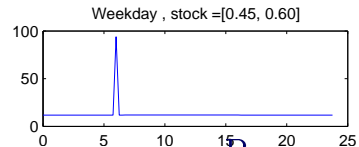
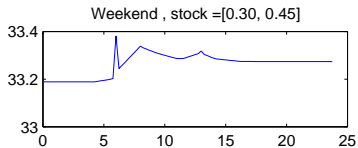
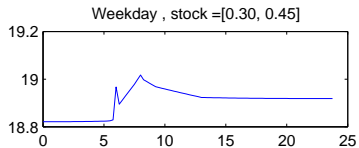
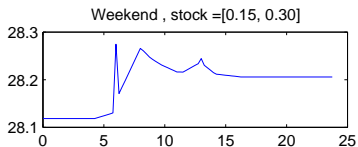
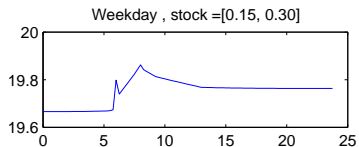
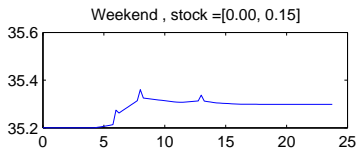
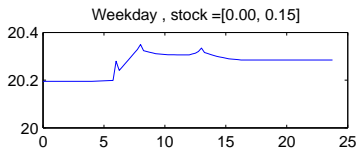
Средняя дисперсия цен. 8 базовых ситуаций:



# Функция комфорта. Регуляризация 0-го порядка



# Функция комфорта. Регуляризация 1-го порядка



# Эффективность предсказания действий

Наилучшие параметры:  $\alpha \in [0.001, 0.1, 0.25]$ ,  $\beta = 20$ ,  $\gamma \in \{0.2, 1, 1.25, 1.5, 2\}$

Эффективность регуляризации 0-го и 1-го порядка одинакова

Дисперсия цен	Низкая	Средняя	Высокая
Алгоритм	Средняя абсолютная ошибка, часы		
kNN	0.80	0.95	0.85
SVM regression	0.65	0.88	0.92
XGBoost	0.34	0.58	0.86
PLS regression	0.87	1.13	0.99
OLS linear regression	2.93	2.93	5.72
Random forest	0.59	0.89	0.96
Mean	1.01	1.16	1.14
Ridge regression	0.94	0.93	0.89
Алгоритмы выявления предпочтений			
Parametric Bayesian learning	0.32	0.39	0.83
Non-parametric IRL	0.29	0.53	0.63

# Заключение

- 1 Предсказание на основе выявления предпочтений превосходит стандартные алгоритмы ML
- 2 Эффективность непараметрического алгоритма IRL сравнима с параметрическим байесовским
- 3 Это означает, что выявление предпочтений имеет смысл применять, даже если вид функции полезности заранее неизвестен

Дальнейшие исследования:

- 1 Разработать модель и алгоритм для HVAC (Heating, Ventilation, Air Conditioning)
- 2 Вероятностный IRL
- 3 Другие приложения?

Подробности в статье:

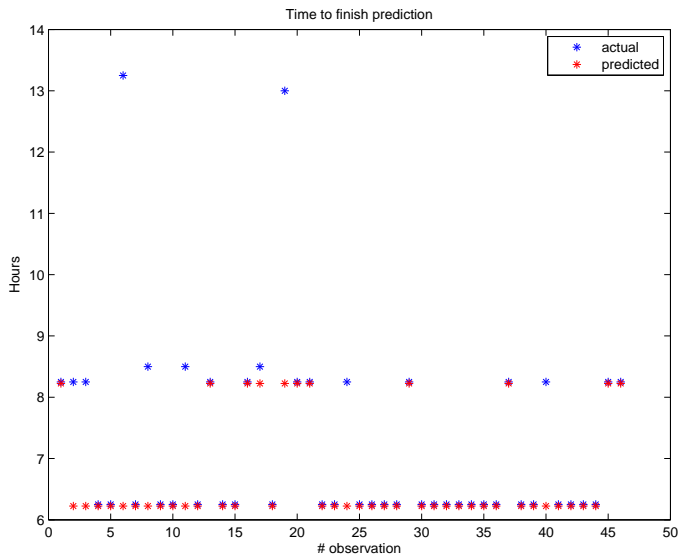
Goubko M.V. et.al. Bayesian Learning of Consumer Preferences for Residential Demand Response. IFAC-PapersOnLine, 2016, Vol. 49, Issue 32, P.24-29

Спасибо за внимание!

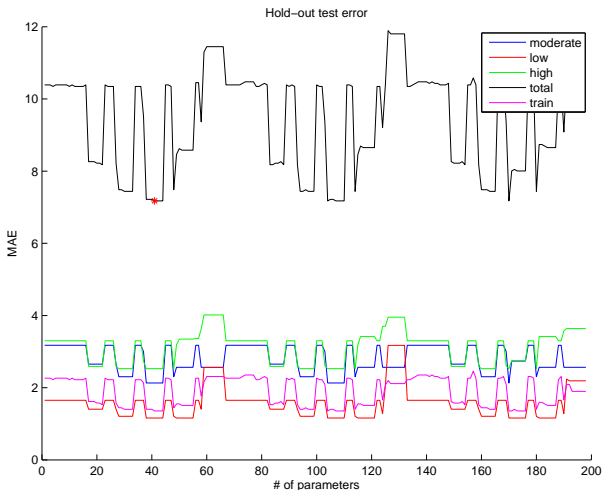
Николай Базенков  
[n.bazenzov@yandex.ru](mailto:n.bazenzov@yandex.ru)

# Actual and predicted finish times

Learning on 80% dataset, the prediction – 20% test dataset:



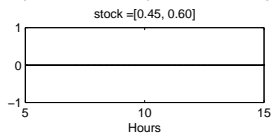
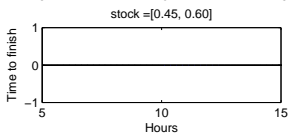
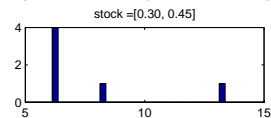
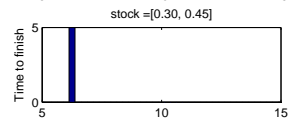
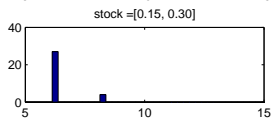
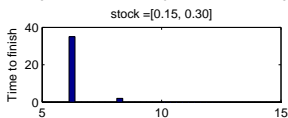
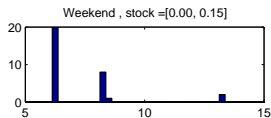
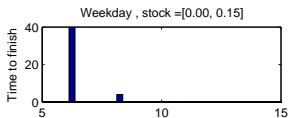
## Hyperparameter search – hold-out



The best combination, independent on regularization:  
 $\alpha \in [0.001, 0.1, 0.25]$ ,  $\beta = 20$ ,  $\gamma \in \{0.2, 1, 1.25, 1.5, 2\}$



## Data - low variance



## Data - high variance

