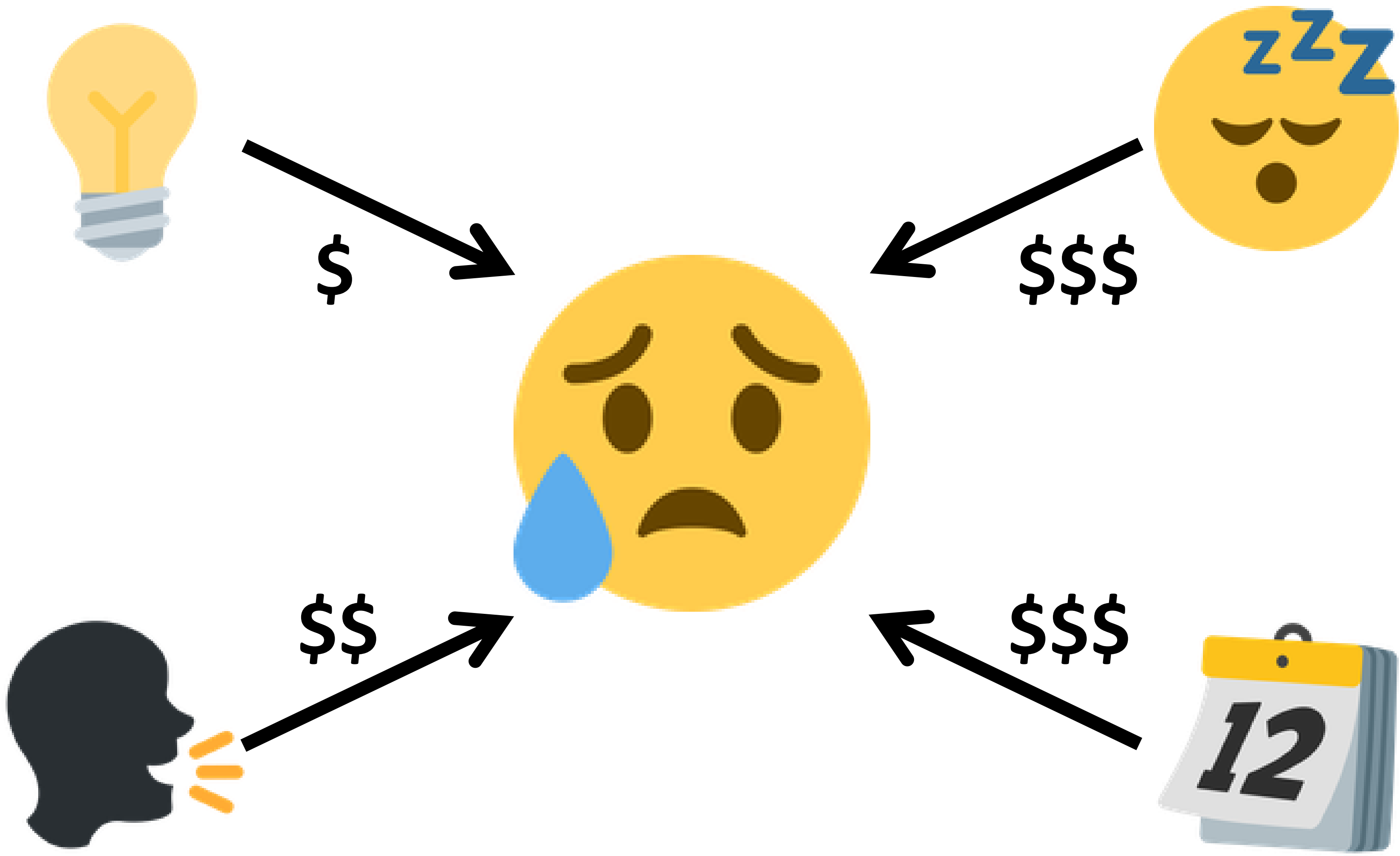
Test Time Feature Ordering with FOCUS: Interactive Predictions with Minimal User Cost

Kirstin Early, Stephen Fienberg, and Jennifer Mankoff
Carnegie Mellon University

Predictive algorithms are a critical part of the ubiquitous computing vision

- Taking appropriate action for users
- "Appropriate": Depends on the situation
  - Often not explicitly specified
  - Needs to be predicted from features
- Supervised learning
  - Training time: Features and labels known
  - Test time: Have new set of features, want to predict label

Example: Predicting a user's stress level



What is test time feature ordering?

- **Test time feature ordering**
  - **Want to make a prediction on a new test point**
  - **Can acquire additional features, at cost**
  - **Order features to acquire**
- *Not* active learning
  - Active learning: *Labels* are costly to obtain at *training time*
  - Our problem: *Features* are costly to obtain at *test time*

Outline

- Problem definition
- Related work
- Our framework
- Validation
  - Energy estimates for prospective tenants
  - Stress prediction for students
  - Device identification for mobile interaction
- Summary

Outline

- Problem definition
- Related work
- Our framework
- Validation
  - Energy estimates for prospective tenants
  - Stress prediction for students
  - Device identification for mobile interaction
- Summary

Past work has taken a variety of approaches for a variety of applications

Paper	Prediction	Cost	Test-Time	Domain
FOCUS	C,R	FS	C/O/#	Health/IoT/Energy
He 2012	C	FS		—
He 2013	C			NLP
Strubell 2015	C		#	NLP
Shi 2015	C			Structured prediction
Weiss 2013	C	FS	#	Structured prediction
Xu 2014	C	FS		LTR, MNIST
Samadi 2015	C	FS	O/#	Knowledge-on-Demand
Pattuk 2015	C	FS	O/#	Information disclosure
Trapeznikov 2013	C	FS	#	—
Golbandi 2011, Sun 2013	C			Recommender Systems

Outline

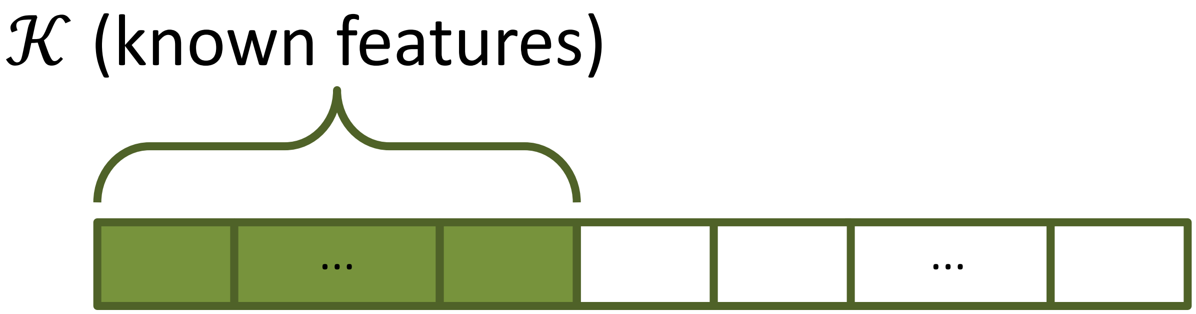
- Problem definition
- Related work
- Our framework
- Validation
  - Energy estimates for prospective tenants
  - Stress prediction for students
  - Device identification for mobile interaction
- Summary

FOCUS

- **F**eature
- **O**rdering with
- **C**ost and
- **U**ncertainty
- **S**core

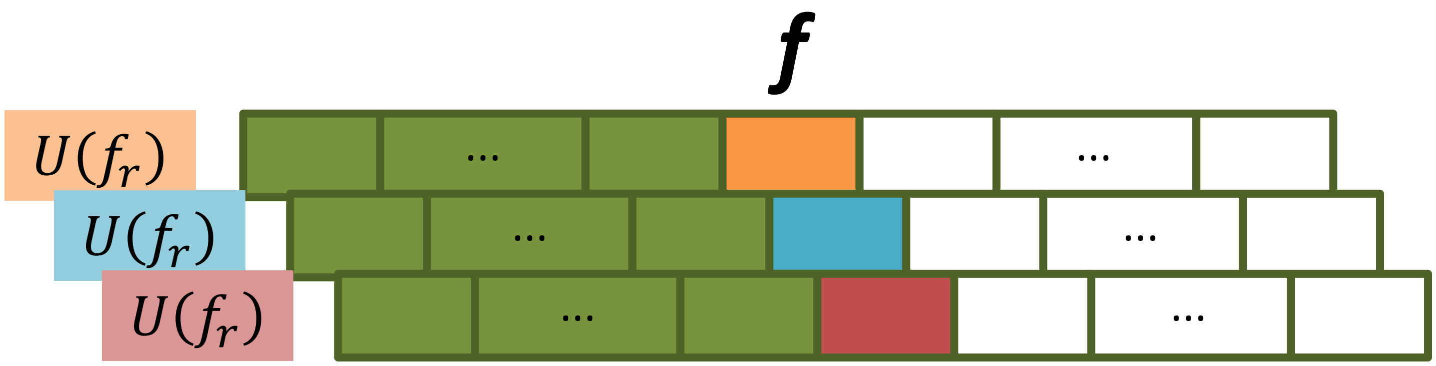
FOCUS iteratively selects features at test time, trading off utility with cost

- Set up: We know a subset of features (green); the rest are unknown (white)
- Goal: Choose an unknown feature to acquire



FOCUS iteratively selects features at test time, trading off utility with cost

- Step 1: Calculate *utility* of each feature *f* we might acquire next
- Utility depends on the actual value of feature, but we can calculate the *expected utility* of *f*



$$\mathbb{E}[U(f)] = \sum_{r \in R} p(f = r) U(f_r)$$

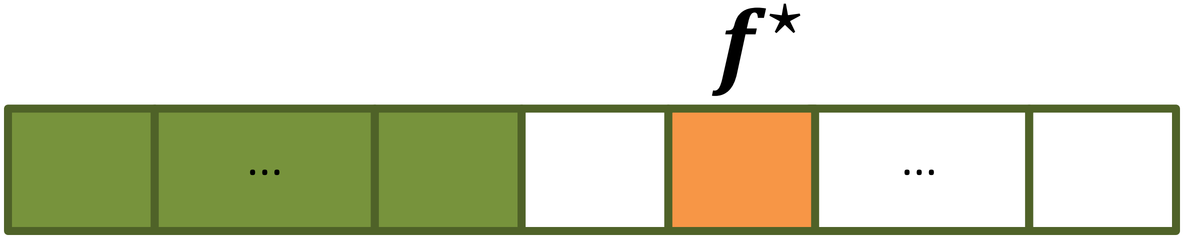
FOCUS iteratively selects features at test time, trading off utility with cost

- Step 2: Combine expected feature utility (calculated in Step 1) with feature cost
- The optimal feature to acquire achieves high expected utility and low cost:

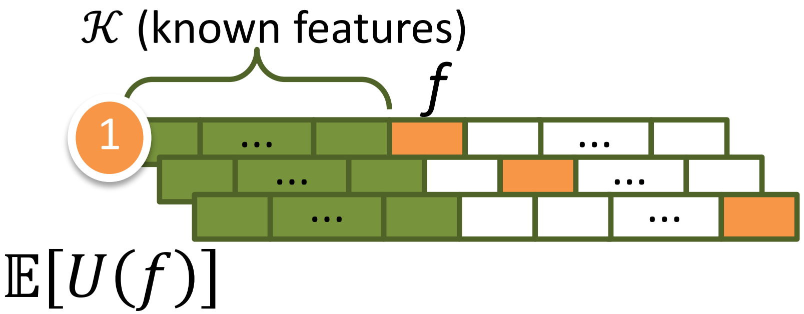
$$f^\star \leftarrow \underset{f \notin \mathcal{K}}{\operatorname{argmin}} \left(-\mathbb{E}[U(f)] + \lambda c_f \right)$$

FOCUS iteratively selects features at test time, trading off utility with cost

- Step 3: Acquire the optimal feature $f^\star$



FOCUS iteratively selects features at test time, trading off utility with cost



2 Optimize to find best combination of feature value and cost, using

$$f^\star \leftarrow \operatorname*{argmin}_{f \notin \mathcal{K}} \left(-\mathbb{E}[U(f)] + \lambda c_f \right)$$

Metrics for feature utility and cost depend on the application

- Example utility metrics
  - Prediction certainty
  - Prediction error
- Example cost metrics
  - Difficulty of acquiring feature
  - Computation time to get feature
  - Battery drain to take sensor measurement

Outline

- Problem definition
- Related work
- Our framework
- Validation
  - Energy estimates for prospective tenants
  - Stress prediction for students
  - Device identification for mobile interaction
- Summary

We compare FOCUS to a fixed-order baseline

- Fixed-order baseline acquires features in the same order, for all test instances
- “Fixed selection”
  - Acquires features in the order of forward selection on the training data

We measure performance with prediction certainty, error, and cost

- Prediction certainty
  - Our utility metric
- Prediction error
  - Regression: Mean absolute error
  - Classification: 0-1 loss
- Cost
  - Sum of all costs of features acquired

Our three validation applications illustrate different aspects of FOCUS

- Energy estimates for prospective tenants
  - Validates: **Feasibility** of FOCUS
  - Algorithm: **Regression**
  - Cost: **User** burden

Our three validation applications illustrate different aspects of FOCUS

- Energy estimates for prospective tenants
  - Feasibility | Regression | User burden
- Stress prediction in college students
  - Validates: **Context-dependent costs**
  - Algorithm: **Regression**
  - Cost: **Battery** drain, **user** burden

Our three validation applications illustrate different aspects of FOCUS

- Energy estimates for prospective tenants
  - Feasibility | Regression | User burden
- Stress prediction in college students
  - Contextual costs | Regression | User/Battery
- Device identification for mobile interactions
  - Validates: **Generalizability** to classification
  - Algorithm: **Classification**
  - Cost: Computation **time**, **user** burden

Our three validation applications illustrate different aspects of FOCUS

- Energy estimates for prospective tenants
  - Feasibility | Regression | User burden
- Stress prediction in college students
  - Contextual costs | Regression | User/Battery
- Device identification for mobile interactions
  - Generalizability | Classification | User/Time

Outline

- Problem definition
- Related work
- Our framework
- Validation
  - Energy estimates for prospective tenants
  - Stress prediction for students
  - Device identification for mobile interaction
- Summary

Selecting energy-efficient homes is important for renters

- But this information is hard to find pre-lease signing
- We can lower user burden by
  - Learning the relationship between household features and energy consumption
  - Using FOCUS to strategically select features that are needed to make a confident prediction for a new household

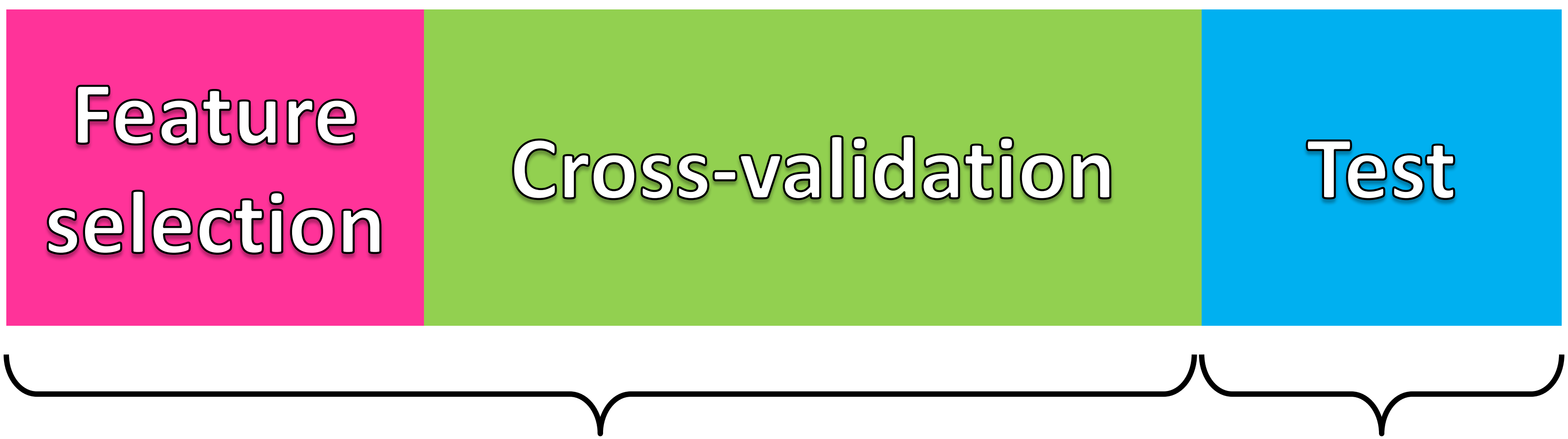
Residential Energy Consumption Survey (RECS)

- Data
  - 12,083 households across U.S.
  - ~500 features of each home
  - Annual energy consumption
- We restrict our use of RECS to electricity prediction in a single climate zone:
  a subset of 2470 homes

Feature cost reflects the effort for a prospective tenant to find an answer

	Meaning	Example feature from RECS
0	Extractable from listing	Number of bedrooms
1	User probably already knows	If someone stays home during the day
2	Might have to check current home	Size of TV
3	Could maybe find in rental listing, or call for easy answer	Washing machine in home
4	Look up online	Year housing unit was built
5	Easily visible during visit	High ceilings
6	Requires effort to find out during visit	Type of glass in windows
7	Requires visit + looking something up	Age of heating equipment

We separate RECS into feature selection, cross-validation, and test sets

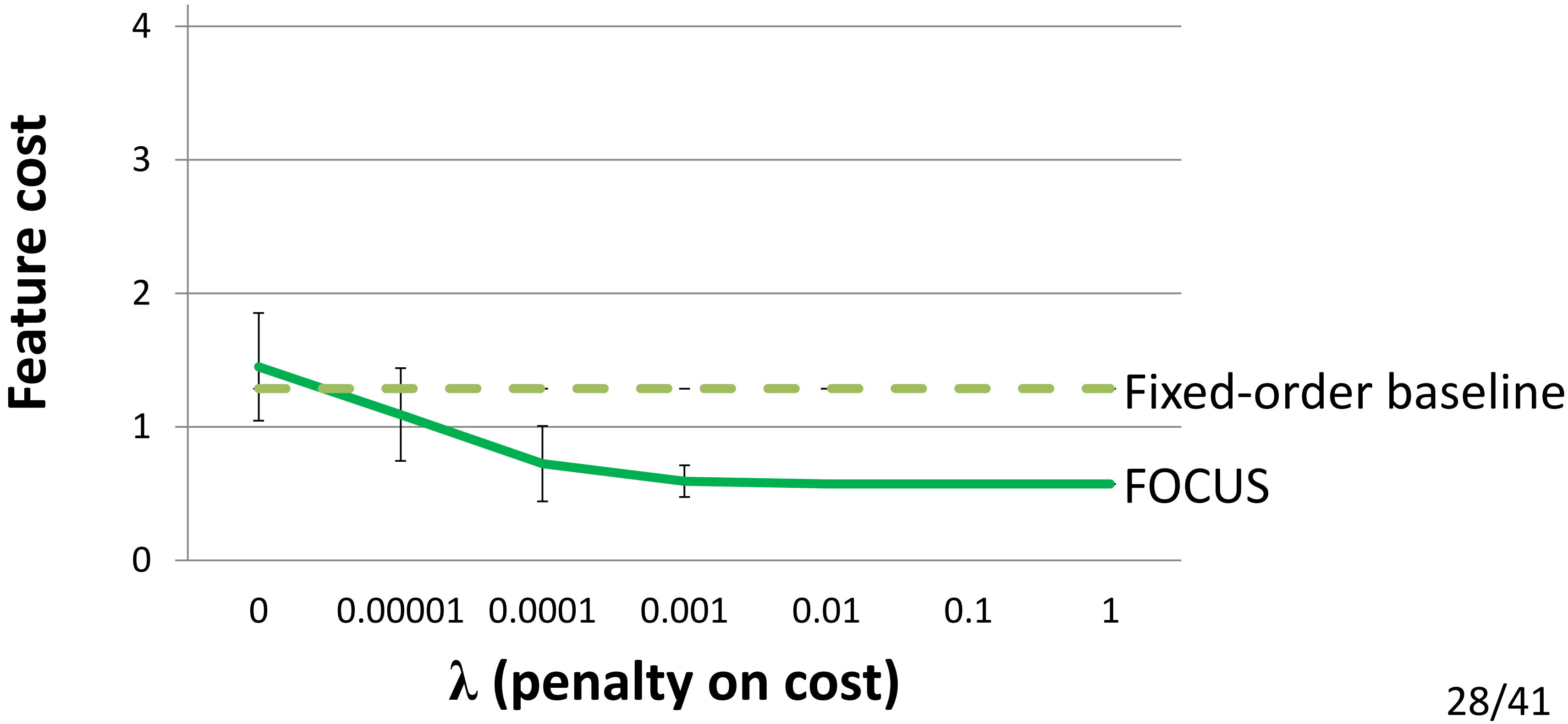


Training set (90%)

Test set (10%)

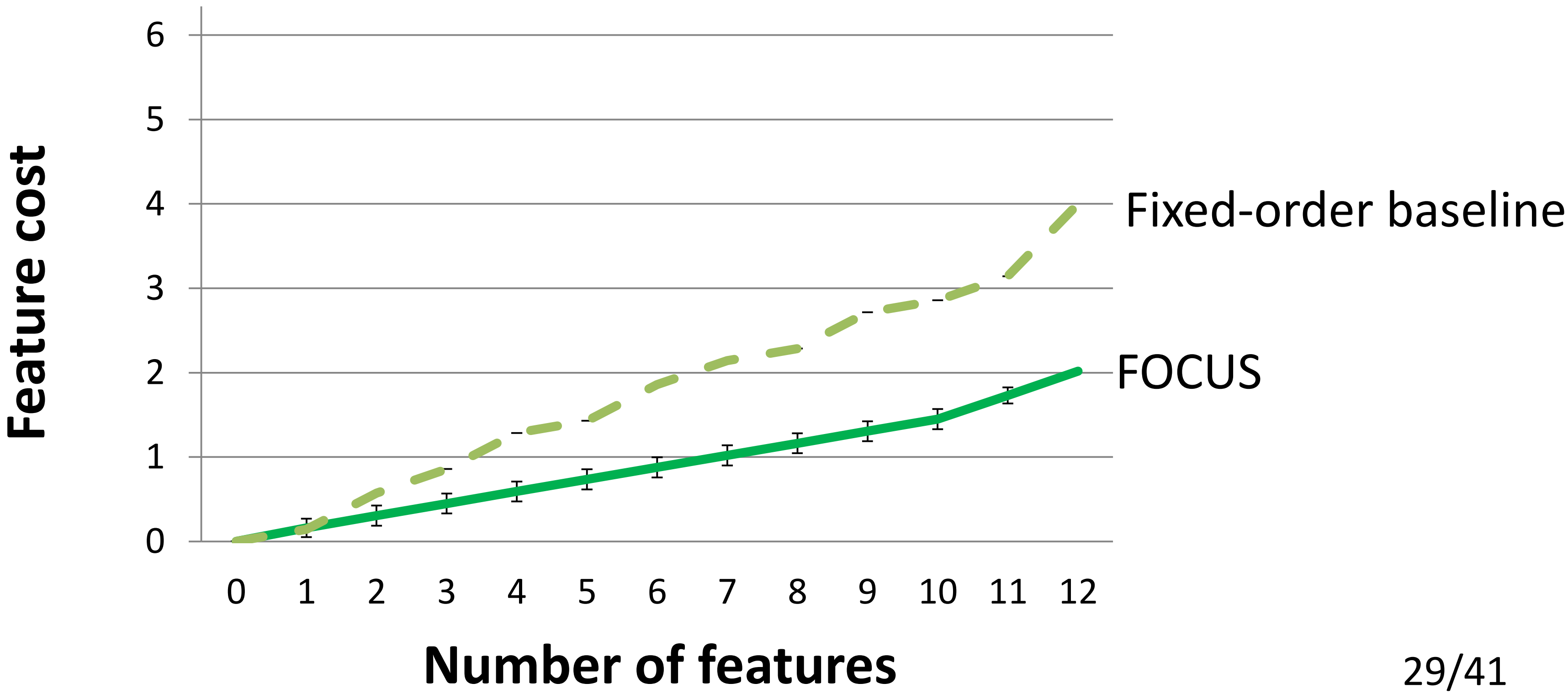
FOCUS yields lower costs than and similar prediction quality as baseline

Cost savings as λ grows, given n=4 features



FOCUS yields lower costs than and similar prediction quality as baseline

Cost increase as additional features are added for λ=.001



Outline

- Problem definition
- Related work
- Our framework
- Validation
  - Energy estimates for prospective tenants
  - Stress prediction for students
  - Device identification for mobile interaction
- Summary

Knowing a user's stress level in the future supports helpful intervention

- We can predict stress reports from data such as demographic information, sleep habits, deadlines, and sensor data
- It is costly to obtain these data
  - Interrupting user to ask question
  - Drain battery to turn on sensor

StudentLife dataset

- Wang *et al*. 2014
- Smartphone data from 48 graduate and undergraduate students over a 10-week academic term
  - Sensor data: Audio, accelerometer, light, *etc.*
  - Self-reported data: Sleep, exercise, *etc*.
  - Survey data: Pre and post psychological surveys
- We focus on self-reports of stress
  - 660 individual stress reports

We predict self-reported stress levels from relevant features and context

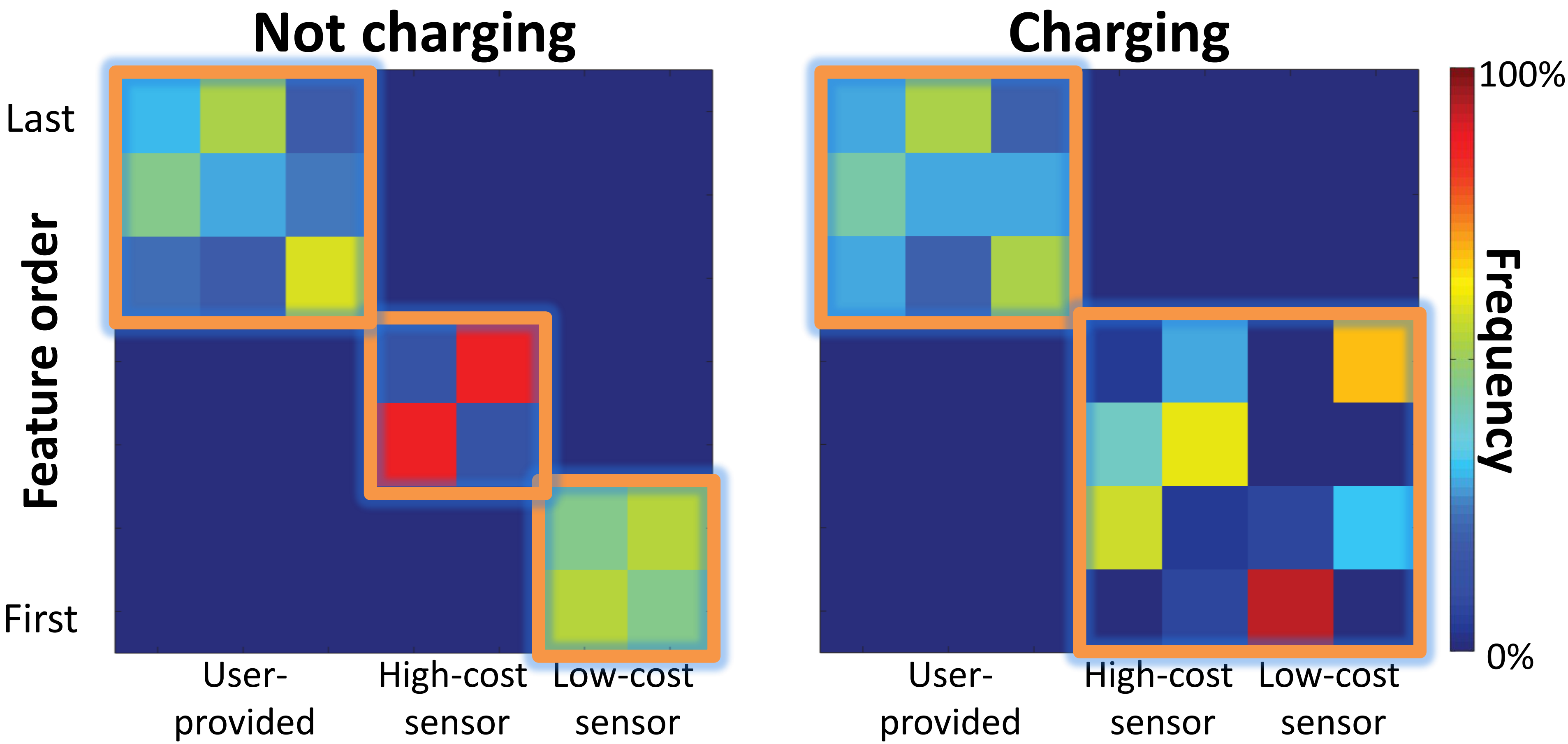
Features associated with stress

- Sleep length
- Exercise length
- Time to next deadline
- Number of upcoming deadlines

Current context

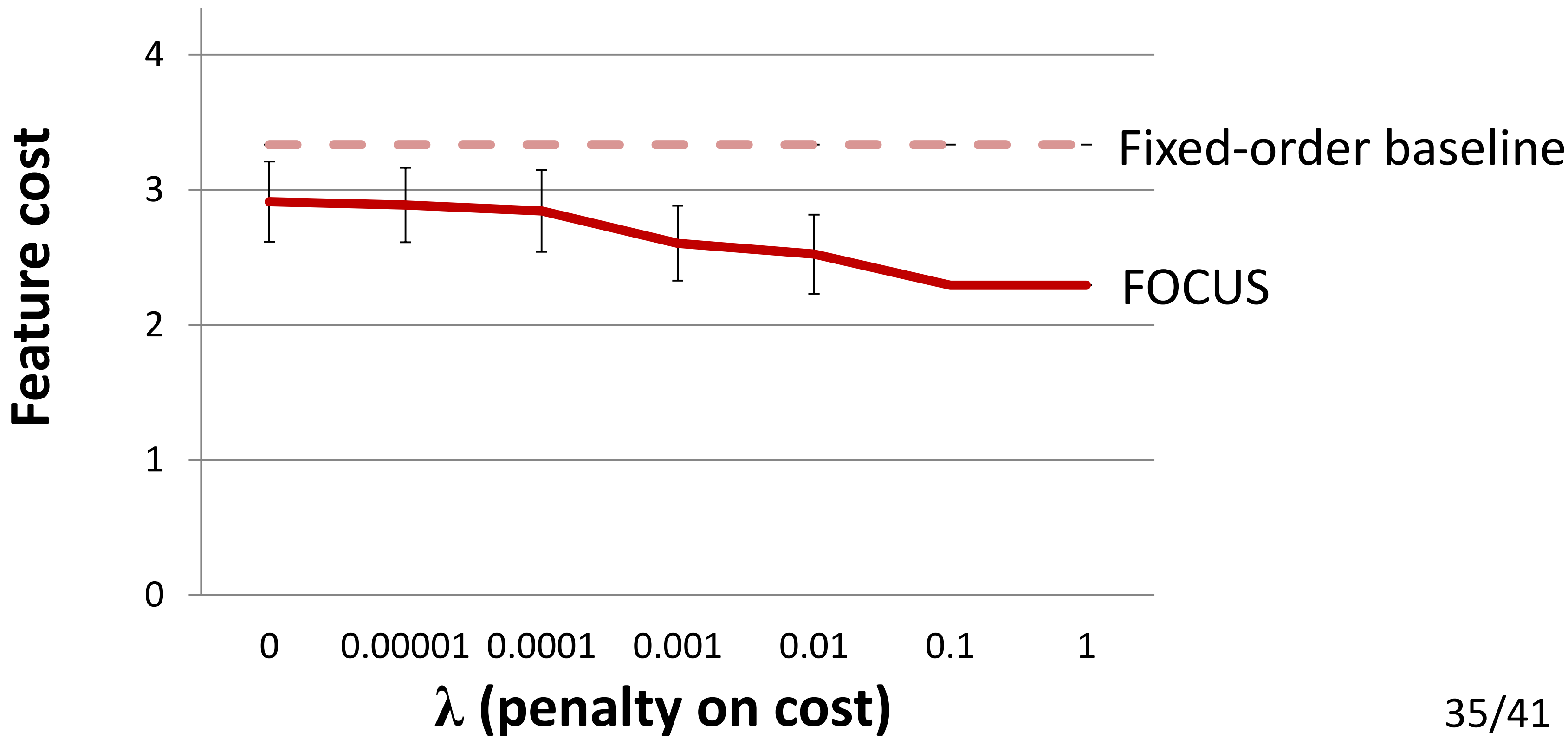
- Time of day (free)
- Activity (high-battery)
- Audio (high-battery)
- Light (low-battery)
- Phone charging (low-battery)

Features with context-dependent costs are selected differently in context



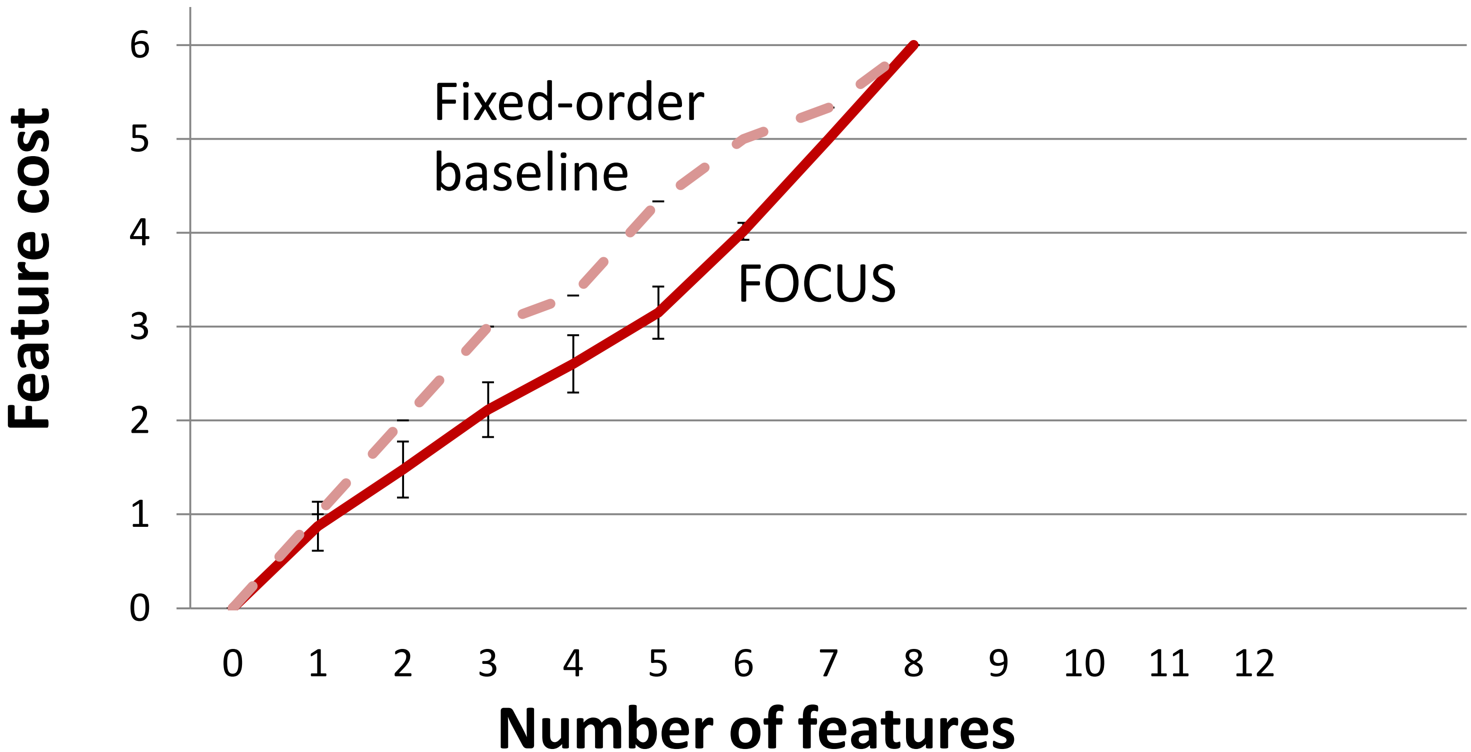
FOCUS yields lower costs than and similar prediction quality as baseline

Cost savings as λ grows, given n=4 features



FOCUS yields lower costs than and similar prediction quality as baseline

Cost increase as additional features are added for λ=.001



Outline

- Problem definition
- Related work
- Our framework
- Validation
  - Energy estimates for prospective tenants
  - Stress prediction for students
  - Device identification for mobile interaction
    - Snap-To-It: Adrian de Freitas *et al*. CHI 2016
- Summary

Outline

- Problem definition
- Related work
- Our framework
- Validation
  - Energy estimates for prospective tenants
  - Stress prediction for students
  - Device identification for mobile interaction
- Summary

Summary

- *Goal*: Make predictions with limited, costly information
- *Solution*: Strategically choose which features to acquire at test time, trading off increases in prediction quality with feature cost
- *Result*: FOCUS predictions are of similar quality as fixed-order baselines, but with cost savings of up to 45%

Limitations and future work

- More nuance for
  - Feature cost definitions
  - Context-dependent cost scenarios
- How to choose tradeoff parameter λ
- Time dependency of feature acquisition

Test Time Feature Ordering with FOCUS: Interactive Predictions with Minimal Cost

Kirstin Early
kearly@cs.cmu.edu
www.cs.cmu.edu/~kearly

References

Adrian de Freitas, Michael Nebeling, *et al*. 2016. Snap-To-It: A user-inspired platform for opportunistic device interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 5909-5920).

He He, Hal Daumé III, & Jason Eisner. (2012). Cost-sensitive dynamic feature selection. In *ICML Inferning Workshop*.

He He, Hal Daumé III, & Jason Eisner. (2013). Dynamic feature selection for dependency parsing. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing* (pp. 1455-1464).

David G. Lowe. (1999). Object recognition from local scale-invariant features. In *Proceedings of the 7th IEEE International Conference on Computer Vision* (pp. 1150-1157).

Erman Pattuk, Murat Kantarcioglu, Huseyin Ulusoy, & Bradley Malin. (2015). Privacy-aware dynamic feature selection. In *31st IEEE International Conference on Data Engineering* (pp. 78-88). IEEE.

Mehdi Samadi, Partha Talukdar, Manuela Veloso, & Tom Mitchell. (2015). AskWorld: Budget-sensitive query evaluation for knowledge-on-demand. In *Proceedings of the 24th International Conference on Artificial Intelligence* (pp. 837-843).

Tianlin Shi, Jacob Steinhardt, & Percy Liang. (2015). Learning where to sample in structured prediction. In *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics (AISTATS)* (pp. 875-884).

Emma Strubell, Luke Vilnis, Kate Silverstein, & Andrew McCallum. (2015). Learning dynamic feature selection for fast sequential prediction. *arXiv preprint arXiv:1505.06169*.

U.S. Energy Information Administration. (2009). Residential Energy Consumption Survey 2009. www.eia.gov/consumption/residential/data/2009/.

Rui Wang, Fanglin Chen, *et al*. (2014). StudentLife: Assessing mental health, academic performance, and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 3-14).

David J. Weiss & Ben Taskar. (2013). Learning adaptive value of information for structured prediction. In *Advances in Neural Information Processing Systems* (pp. 953-961).

Zhixiang (Eddie) Xu, Matt J. Kusner, Killian Q. Weinberger, Minmin Chen, & Olivier Chapelle. (2014). Classifier cascades and trees for minimizing feature evaluation cost. *The Journal of Machine Learning Research*, *15*(1), 2113-2144.