

Guidelines for Intelligent Interfaces

Daniel Weld
University of Washington

Acknowledgements

- Krzysztof Gajos
- Corin Anderson
- Mary Czerwinski
- Pedro Domingos
- Oren Etzioni
- Raphael Hoffman
- Tessa Lau
- Desney Tan
- Steve Wolfman
- UW AI Group
- DARPA, NSF, ONR, WRF, Microsoft Research



Early Adaptation: Mitchell, Maes

- Predict: Email message priorities
Meeting locations, durations

2:00	I* Huang	I- Zabowski	✓	
	IWeh5309	IWeh5309	-----	
2:30	I- Dent		I* Dent	
	IWeh5309		IWeh5309	
3:00	I- Theo-Groul	I- Chalasani	I- Many	
	IWeh4605	IWeh5309	IWeh4605	
3:30	IResearch	I* Jourdan	DI Course	I* Jourdan
		IWeh5309		IWeh5309
4:00	✓		✓	I- Cheng
	-----		✓	IWeh5309
4:30	I- Yamanouchi		✓	IResearch
	IWeh5309		✓	
5:00	IResearch			
		4-2	4-3	4-4

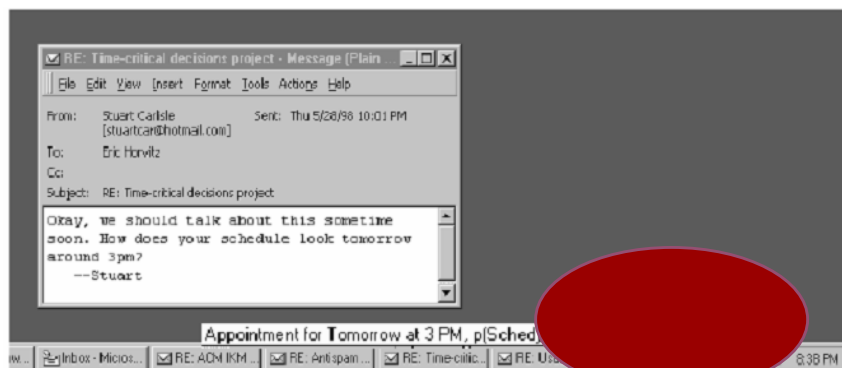
- Principle 1: *Defaults* minimize cost of errors
- Principle 2: *Allow users to adjust thresholds*

14-Mar-19

Daniel S. Weld / Univ. Washington

21

Adaptation in Lookout: Horvitz



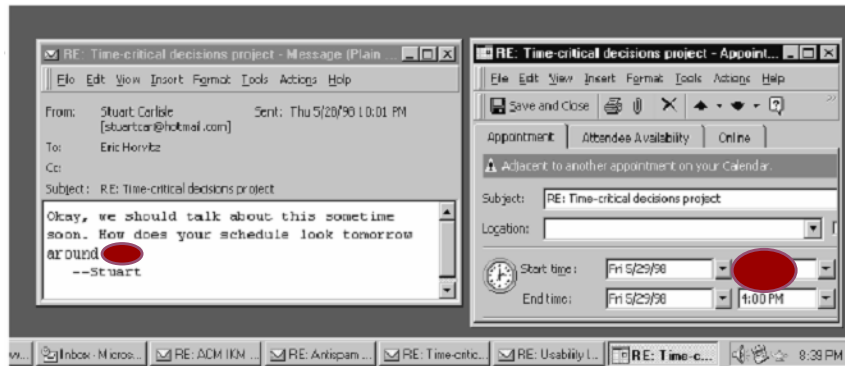
14-Mar-19

Daniel S. Weld / Univ. Washington

Adapted from Horvitz

22

Adaptation in Lookout: Horvitz



14-Mar-19

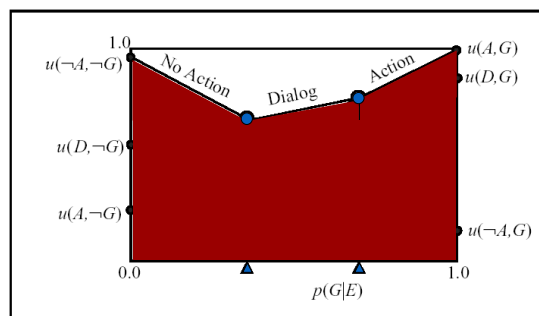
Daniel S. Weld / Univ. Washington

23

Resulting Principles

[Horvitz CHI-99]

- Decision-Theoretic Framework
 - Graceful degradation of service precision
 - Use dialogs to disambiguate (Considering cost of user time, attention)



14-Mar-19

Daniel S. Weld / Univ. Washington

Adapted from Horvitz

24

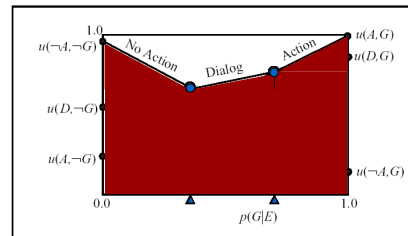
Horvitz <-> POMDP?

- What's Shared?

- Policy mapping from belief state to action
- Idea of maximizing utility

- What's Different?

- No model of state transition
- No lookahead or notion of time
- Greedy policy

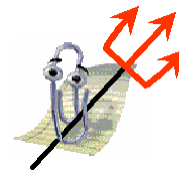


25

Principles About Invocation

Allow efficient invocation, correction & dismissal

Timeouts minimize cost of prediction errors



20 Year Retrospective

- More guidelines
- <https://medium.com/microsoft-design/guidelines-for-human-ai-interaction-9aa1535d72b9>

Human-AI Teams



- Environment gives percept
- AI makes recommendation [+ explanation]
- Human decides whether to
 - Trust AI's advice, or
 - Get more info and decide herself
- Reward based on speed/accuracy

*Over time, human builds trust model
[Rouse & Morris '86, etc.]*

Updates in Human-AI Teams



- Environment gives percept
- AI makes recommendation [+ explanation]
- Human decides whether to
 - Trust AI's advice, or
 - Get more info and decide herself
- Reward based on speed/accuracy



Gagan Bansal

Besa Nushi

Ece Kamar

Walter Lasecki

Eric Horvitz

[Bansal et al. AAAI19]

Many ML Algorithms aren't Stable wrt Updates

Classifier	Dataset	ROC h_1	ROC h_2
LR	Recidivism	0.68	0.72
	Credit Risk	0.72	0.77
	Mortality	0.68	0.77
MLP	Recidivism	0.59	0.73
	Credit Risk	0.70	0.80
	Mortality	0.71	0.84

When trained on more data (same distribution)...

- Updates (h_2) increase **ROC**...

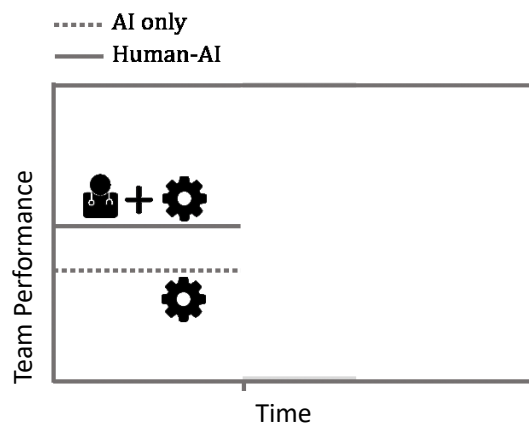
Many ML Algorithms aren't Stable wrt Updates

Classifier	Dataset	ROC h_1	ROC h_2	CS
LR	Recidivism	0.68	0.72	0.74
	Credit Risk	0.72	0.77	0.68
	Mortality	0.68	0.77	0.54
MLP	Recidivism	0.59	0.73	0.62
	Credit Risk	0.70	0.80	0.69
	Mortality	0.71	0.84	0.77

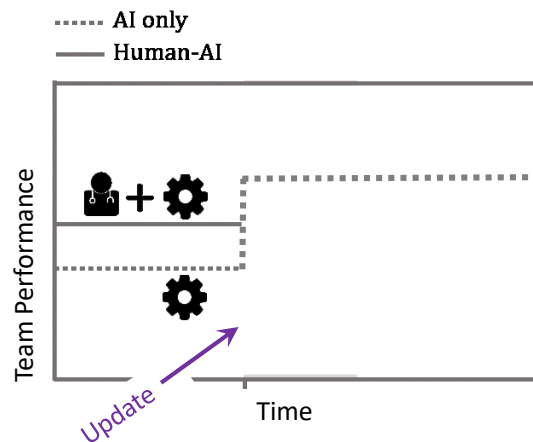
When trained on more data (same distribution)...

- Updates (h_2) increase ROC, $\alpha_{h_1, h_2} = 1 - \frac{\text{count}(h_1 = y, h_2 \neq y)}{\text{count}(h_2 \neq y)}$
- But have low **compatibility score**,

But for Teams, ...



But for Teams, Updates ...



But for Teams, Updates should be Compatible

