User's Feedback in Preference Elicitation

Tereza Siváková

24.6.2022

Contents

Decision Making Runs in Closed Loop

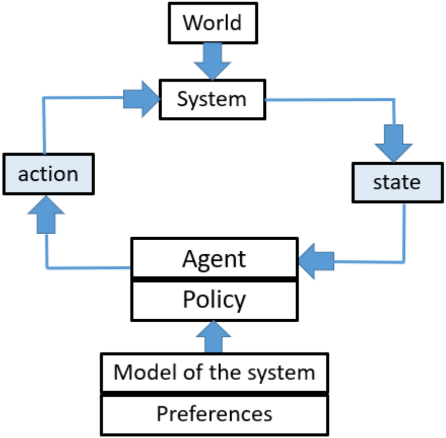


Figure: Schema of the closed loop

...quantification of user's preferences is hard

Fully Probabilistic Design (FPD)

It introduces the ideal distribution of behavior

$$c^i(b) = \prod_{t \in \mathbf{T}} m^i(s_t|a_t, s_{t-1}) r^i(a_t|s_{t-1}),$$

to which the real one

$$c^\pi(b) = \prod_{t \in \mathbf{T}} m(s_t|a_t, s_{t-1}) r(a_t|s_{t-1})$$

tries to get closer. The similarity of these two distributions is measured by Kullback-Leibler divergence

$$D(c^\pi||c^i) = \int_{b \in \mathbf{B}} c^\pi(b) \ln\left(\frac{c^\pi(b)}{c^i(b)}\right).$$

Then the optimal policy is

$$\pi^0 \in \arg\min_{\pi \in \Pi} D(c^\pi||c^i).$$

The selected actions a and observed states s up to the horizon $h \in \mathbb{N}$

$$b = (s_0, a_1, s_1, a_2, \ldots, a_h, s_h),$$

describe the behavior.

The sequence of decision rules

$$\pi = (r(a_1|s_0), r(a_2|s_1), .., r(a_h|s_{h-1})),$$

forms action generating policy.

The sequence of probability densities

$$m = (m(s_1|a_1, s_0), m(s_2|a_2, s_1), .., m(s_h|a_h, s_{h-1}))$$

forms the model of the system.

The Problem: Ideal Distribution c^i is Hard to Specify

The optimal ideal distribution : $c_0^i \in \arg\min_{c^i \in \mathbf{C}^i} \min_{\pi \in \Pi} D(c^\pi || c^i)$.

Solved for

- $$\mathrm{r}^i \in \mathbb{R}^i \equiv \left\{ \mathrm{r}^i : \mathrm{supp}[\mathrm{r}^i] = \mathbb{A} \right\}, \text{it provides exploration} \tag{1}$$
- $(1 - w)$ the probability $(\mathbb{S}^i) + w$ probability $(\mathbb{A}^i)$ should be maximal

Problem

Satisfactory weight $w \in [0, 1]$ is unknown.

Solution: User Judges Behavior for Given w

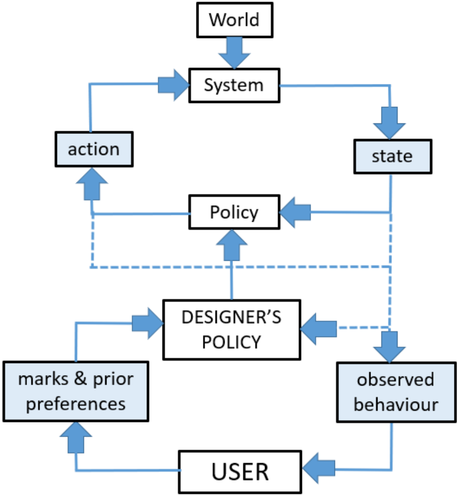


Figure: Schema of the upper level closed loop for PE.

Simulated system

Comparison of number of states and actions for the system 15x7x15, which was created using the "heat equation"

$$y_t = 0.028y_{t-1} + 1.81y_{t-2} - 0.817y_{t-3} + 0.1a_t - 0.16a_{t-1} + 0.05\mathcal{N}(0,1), \quad (2)$$

where a_t is a uniformly selected discrete action from $\mathbb{A} = \{1, \ldots, 7\}$. Thousand of sample Y_t were discretized and affine mapped on $\mathbb{S} = \{1, \ldots, 15\}$. The occurrence of triplets was mapped on simulated transition probabilities.

Price paid for individual actions

action	1	2	3	4	5	6	7
price	3	2	1	0	1	2	3

Price paid for individual states

state	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
price	4	3	3	2	2	1	1	0	1	1	2	2	3	3	4

Behavior wanted by the user:

- Preferred set of states $\mathbb{S}^i = \{8\}$ and of actions $\mathbb{A}^i = \{4\}$.

Marking by the user:

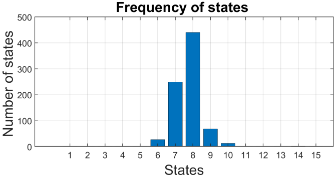
- Every 10 steps the state and action sequences were shown
- The user marked the sequences (school marks $1, \ldots, 5$)

Upper level decision making:

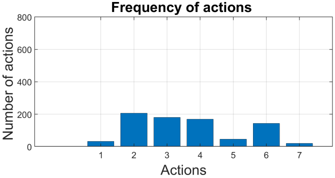
- The upper model related weight w and internal ν to the mark
- The upper feedback choose w, ν to get the best mark
- The simulation continued with the improved w and ν.

Experiment 1.

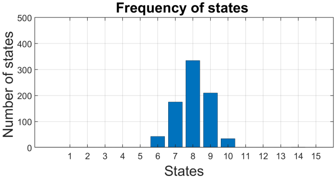
Simulations without the user.



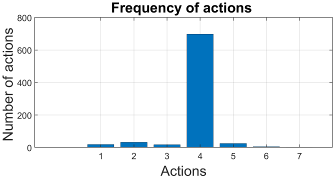
(1) States for $\mathbb{S}^i = \{8\}, \mathbb{A}^i = \mathbb{A}$ $w = 0$



(2) Actions for $\mathbb{S}^i = \{8\}, \mathbb{A}^i = \mathbb{A}$ $w = 0$



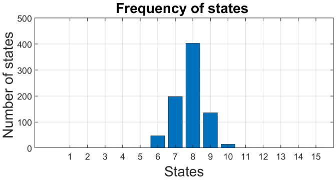
(3) States for $\mathbb{S}^i = \{8\}, \mathbb{A}^i = \{4\}$ $w = 0.3$



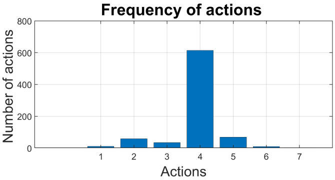
(4) Actions for $\mathbb{S}^i = \{8\}, \mathbb{A}^i = \{4\}$ $w = 0.3$

Experiment 2.

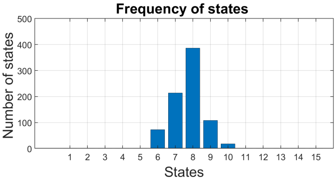
Preferred $\mathbb{S}^i = \{8\}$, $\mathbb{A}^i = \{4\}$ both weight w and ν were tuned via the user's marks.



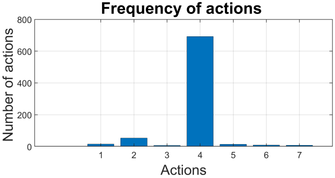
(5) States for the 1st user



(6) Actions for the 1st user



(7) States for the 2nd user

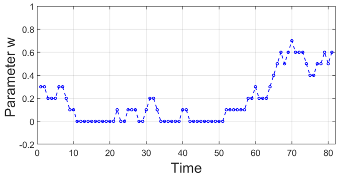


(8) Actions for the 2nd user

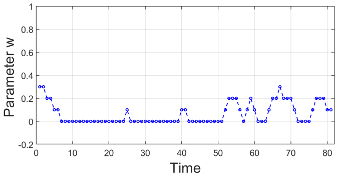
Comparison

The price paid for actions of all experiments.

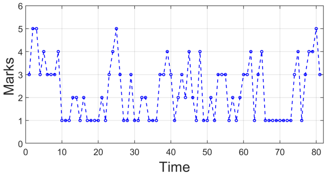
Parameters	Actions price	States price	Total price	Number of preferred action	Number of preferred state
$w = 0, \ \nu = 1$	1086	370	1456	170	440
$w = 0.3, \ \nu = 1$	181	475	656	698	335
1^{st} user	281	403	684	614	403
2^{nd} user	219	420	639	692	386



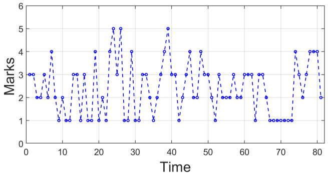
(9) Parameter w in time for the 1^{st} user



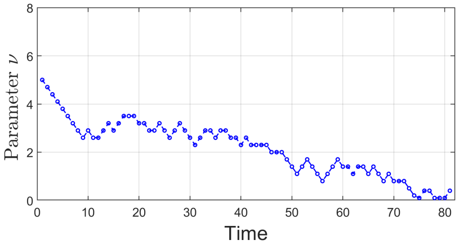
(10) Parameter w in time for the 2^{nd} user



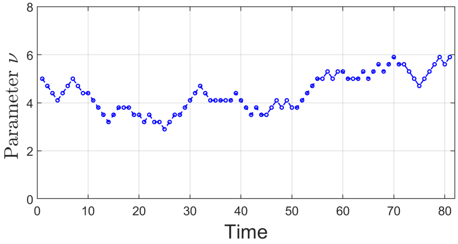
(11) Evolution of marks of 1^{st} user



(12) Evolution of marks of 2^{nd} user



(13) Parameter ν in time for 1st user



(14) Parameter ν in time for 2nd user

Conclusion

- The DM with and without user‘s control was compared.
- Preferences elicitation is an important and hard task.
- Open: missing more realistic systems with larger dimension
- Adding more free parameters to the upper level closed-loop, for example extension of the sets of preferred states and actions.
- Experiments with more users.
- Continuous states/actions, fight with the dimensionality curse

Thank you for your attention!

References I

J. Drummond and C. Boutilier.
Preference elicitation and interview minimization in stable matchings.
In *Proc. of 28th AAAI Conf. on AI*, pages 645 – 653, 2014.

I. Garcia, S. Pajares, L. Sebastia, and E. Onaindia.
Preference elicitation techniques for group recommender systems.
Information Sciences, 189:155–175, 2012.

Shengbo Guo and Scott Sanner.
Real-time multiattribute Bayesian preference elicitation with pairwise comparison queries.
In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pages 289–296, Chia Laguna Resort, Sardinia, Italy, 2010. JMLR Workshop and Conference Proceedings.

M. Kárný and T.V. Guy.
Preference elicitation within framework of fully probabilistic design of decision strategies.
In *IFAC Int. Workshop on Adaptive and Learning Control Systems*, volume 52, pages 239–244, 2019.

M. Kárný and T. Siváková.
Agent's feedback in preference elicitation.
In *2021 20th International Conference on Ubiquitous Computing and Communications (IUCC/CIT/DSCI/SmartCNS)*, pages 421–429, 2021.

L. Naamani-Dery, M. Kalech, L. Rokach, and B. Shapira.
Reducing preference elicitation in group decision making.
Expert Systems with Applications: An International Journal, 61:246–261, 2016.

V. Peterka.
Bayesian system identification.
In P. Eykhoff, editor, *Trends and Progress in System Identification*, pages 239–304. Perg. Press, 1981.

References II

M.L. Puterman.
Markov Decision Processes: Discrete Stochastic Dynamic Programming.
Wiley, 2005.

P. Viappiani and C. Boutilier.
Optimal Bayesian recommendation sets and myopically optimal choice query sets.
Advances in Neural Information Processing Systems 23, pages 2352–2360, 2010.