An application of preference learning to the selection of blockchain validators

Grzegorz Miebs[a], Jonas Gehrlein[b], Matteo Brunelli[c], Milosz Kadzinski[a]

[a]Poznan University of Technology,Poland
[b]Web3 Foundation, Switzerland
[c]University of Trento, Italy

Introduction

1. We apply an active learning algorithm to solve a MCDM-problem in the blockchain space.
2. The decision problem, with hundreds of alternatives, is faced by hundreds of people daily.
3. We developed an algorithmic solution and tested it with real decision makers in an experiment.

The decision problem (simplified)

1. There are two key players in this situation:
   1. **Validators** maintain the blockchain and guarantee that the transactions that are submitted to it are valid.
   2. "**Nominators** secure the Relay Chain by selecting good validators and staking DOT".
2. Nominators can select up to 16 Validators from a huge list of alternatives
3. There are fundamental trade-offs to consider between security, profitability, and decentralization
4. Each nominator might have different preferences on those criteria.
5. An active involvement of nominators is required to filter out the good from the bad validators.
6. There is too much data for nominators to properly process
7. Underlying set of validators changes frequently and your selection today might not be optimal tomorrow.

The decision problem (simplified)

1. There are two key players in this situation:
    1. **Validators** maintain the blockchain and guarantee that the transactions that are submitted to it are valid.
    2. “**Nominators** secure the Relay Chain by selecting good validators and staking DOT”.
2. Nominators can select up to 16 Validators from a huge list of alternatives
3. There are fundamental trade-offs to consider between security, profitability, and decentralization
4. Each nominator might have different preferences on those criteria.
5. An active involvement of nominators is required to filter out the good from the bad validators.
6. There is too much data for nominators to properly process
7. Underlying set of validators changes frequently and your selection today might not be optimal tomorrow.

$\Rightarrow$ Need for simplification

Additive value functions

If a set of attributes is mutual preferential independent, then the value of an alternative a is given by an **additive value function**

$$U(a) = \sum_{j=1}^{n} u_j(g_i(a))$$

which can represent the preferences such that

$$U(a) \geq U(b) \quad \Leftrightarrow \quad a \succsim b \quad \forall a, b \in A$$

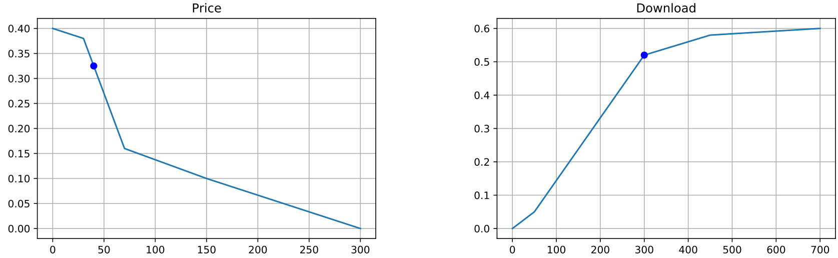
Additive value functions

Consider a connection with the following two attributes

1. price: 40E
2. download: 300 Mbp/s

Then if we know the attribute value functions u_{price} and u_{download}

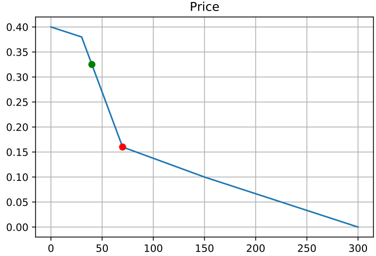
$$U(a) = u_{\text{price}}(40) + u_{\text{download}}(300) = 0.325 + 0.52 = 0.845$$

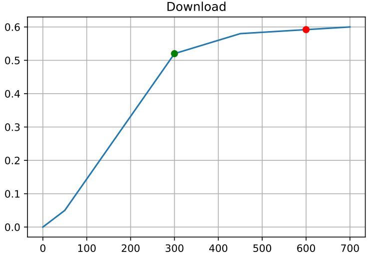


Additive value functions

Consider two alternative internet connections

	price	download
a	40	300
b	70	600





If we use the previous value functions

$$\overbrace{0.325 + 0.52}^{U(a)} > \overbrace{0.16 + 0.592}^{U(b)} \quad \Rightarrow \quad a \succ b$$

The attributes for the choice of validators

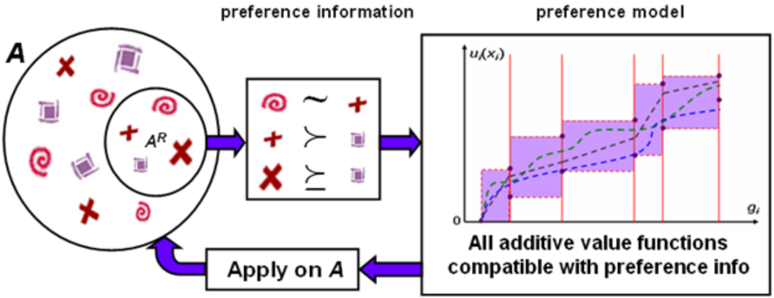
j	Attribute name	Measurement unit	X_j	monotone
1	Commissions	%	$[0, 100]$	$\searrow^*$
2	Self-stake	dots	$[1, 604440]$	$\nearrow$
3	Total stake	dots	$[1942012.9, 5906988.5]$	$\searrow$
4	Era points	number	$\{520, \ldots, 1260\}$	$\nearrow$
5	cluster size	number	$\{1, \ldots, 21\}$	Λ
6	Voters	number	$\{46, \ldots, 3172\}$	Λ

These attributes are:

- unambiguous, comprehensive, direct, operational and understandable
- mutual preferential independent for non-extreme values.
- tangible and quantitative

Preference learning

Siskos, Y., Grigoroudis, E., & Matsatsinis, N. F. (2016). UTA methods. In: *Multiple Criteria Decision Analysis* (pp. 315-362). Springer, New York, NY.



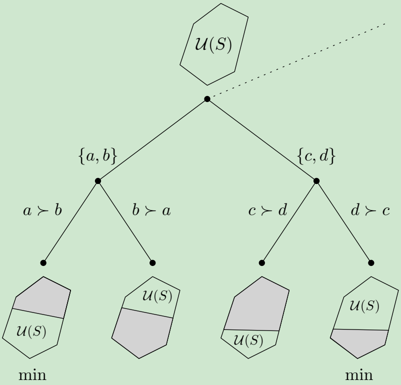
We call $\mathcal{U}(S)$ the set of compatible value functions.

On the next question

We can choose from a large pool of alternatives, but how?

Example

Consider two pairs $\{a, b\}$, $\{c, d\}$

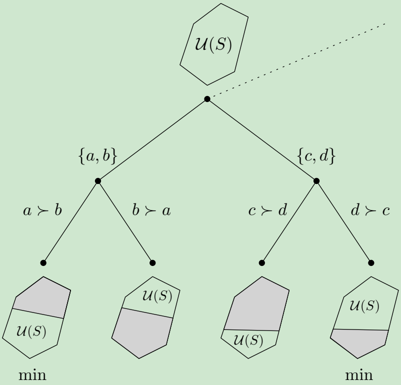


On the next question

We can choose from a large pool of alternatives, but how?

Example

Consider two pairs $\{a, b\}$, $\{c, d\}$



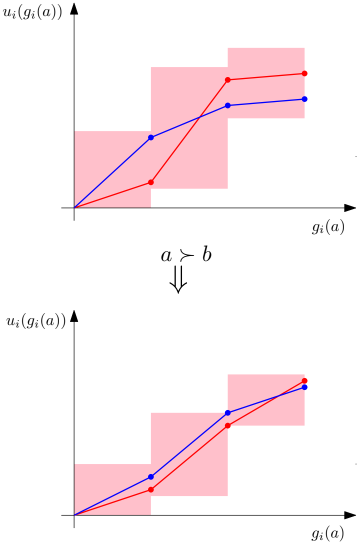
Our approach

We adopt a **conservative** view and we choose the pair $\{a, b\}$. In the worst case we get the most information.

How to estimate the gain of information

An indirect estimation

- Assume $a \succ b$
- Sample $s = 1000$ value functions compatible with $\mathcal{U}(S)$.
- Calculate the Spearman rank correlation between each one of them.
- Aggregate the obtained values of Spearman rank correlation by taking the minimal value.
  - The closer to 1, the more similar the rankings are, and the smaller $\mathcal{U}(S)$ must be.

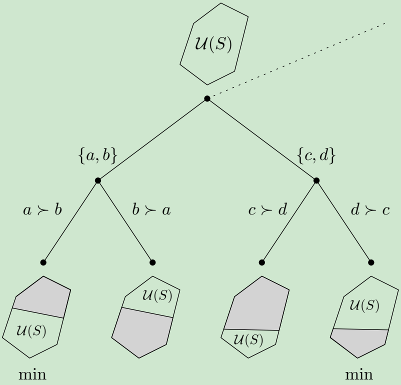


Fictive alternatives

Can we make it even better?

Example

Consider two pairs $\{a, b\}$, $\{c, d\}$

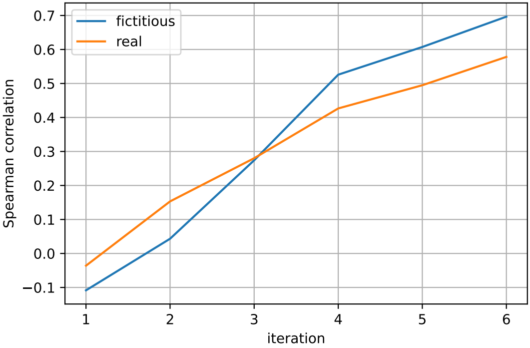


Fictive alternatives

With the introduction of the fictive alternative optimal question can be found

The decreasing size of $\mathcal{U}(S)$

It was estimated by the growth of the Spearman index of the obtained rankings



On the experiment with validators and nominators

An experiment with

- real nominators and validators
- rewards depending on their answers

On the experiment with validators and nominators

An experiment with

- real nominators and validators
- rewards depending on their answers

1. Manual selection

Ask the nominator to choose 7 validators out of a list of more than 200

2. Algorithmic selection

1. Ask the nominator 6 pairwise comparisons
2. estimate his/her value functions
3. select the best 7 validators

3. Comparison

1. Show the union of the manual and algorithmic selections (≤ 14) and ask to select the best 7.
2. Evaluate (questionnaire, return, choice analysis)

Manual selection

Commission (in %)	Self Stake (in DOT)	Total Stake (in DOT)	Avg. Era Points	Cluster Size	Voters	
2.5	10.0	2182626.5	19760.8	15	960	Select
3.0	1.1	2182713.6	20458.04	8	182	Select
100.0	1.0	1947313.6	21134.12	1	8	Select
20.0	1.0	2182611.9	20784.71	6	259	Select
100.0	2.0	1937122.6	19984.31	1	0	Select
2.0	442.3	2182634.2	20728.24	1	1465	Select
100.0	2.0	1999898.2	19259.61	1	2	Select
5.0	604440.2	3430973.7	21340.39	1	1729	Select
100.0	20.0	1884214.6	21877.25	1	51	Select
100.0	1.0	1877784.3	20511.37	11	14	Select
100.0	2.0	1938938.2	20796.47	1	2	Select
100.0	1.0	1883439.4	19082.98	1	5	Select

Algorithm

INSTRUCTIONS

In the following, we present you a choice between two validators and ask you to select which of the two you prefer. We ask you to answer six of those pairwise comparisons. Afterwards, the active learning algorithm selects the most suited validators from all available validators. If you read the instruction and can start please click ready.

SELECTION 4/6

Commission (in %)	Self stake (in dot)	Total stake (in dot)	Avg. era points	Cluster size	Voters	
1.6	284304.2	2042681.1	19553.0	15.0	3011.0	Select
2.7	413274.6	2903988.2	20272.0	3.0	2625.0	Select

Recommendation

Thank you for your answers. These are your recommendations:

Commission (in %)	Self stake (in dot)	Total stake (in dot)	Avg. era points	Cluster size	Voters
0.5	11039.0	2182623.3	24398.82	12.0	1284.0
1.0	624586.0	2182629.8	21241.96	1.0	2330.0
0.0	3.0	2182549.5	25765.71	9.0	240.0
3.0	1050051	2182622.3	22084.31	1.0	1095.0
0.5	1110.0	2182627.3	22807.45	1.0	1972.0
3.0	55400.5	2182659.3	22246.27	1.0	2341.0
3.0	17327.3	2182630.0	21972.44	1.0	744.0

Please take a brief look at the algorithm recommendation. You can then proceed to the next step.

Final choice

Selection	Commission (in %)	Self Stake (in DOT)	Total Stake (in DOT)	Avg. Era Points	Cluster Size	Voters	
B	3.0	105005.1	2182622.3	22084.31	1	1095	Select
A	3.0	1.0	2182562.6	24865.52	8	229	Select
B	3.0	55400.5	2182659.3	22246.27	1	2341	Select
B	3.0	17327.3	2182630.0	21972.44	1	744	Select
AB	0.5	1110.0	2182627.3	22807.45	1	1972	Select
A	0.0	3.0	2182645.6	22690.59	9	2113	Select
B	1.0	624586.0	2182629.8	21241.96	1	2330	Select
AB	0.0	3.0	2182549.5	25765.71	9	240	Select
A	0.5	22.0	2182631.8	22734.12	12	1657	Select
AB	0.5	11039.0	2182623.3	24398.82	12	1284	Select
A	2.5	10.0	2182677.2	24370.0	15	1924	Select

Questionnaire

Are you or have you been a validator on Polkadot or Kusama?*

- Yes
- No

How well does the recommendation suit your preferences?*

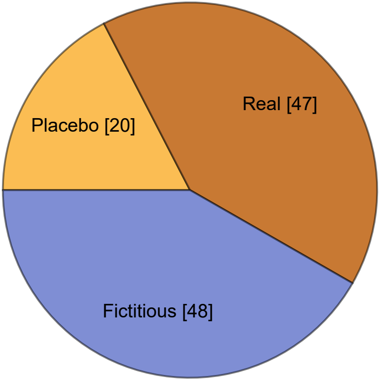
- Very well
- To some extent
- Not very well
- Not at all

How would you rate the difficulty of manually selecting validators (Part A)?*

- Very easy
- Easy
- Medium
- Hard
- Very hard

How would you rate the difficulty of the pairwise comparisons (Part B)?*

The participants in the experiment...
... were splitted into three groups



Results

Choice

	random	fictitious	real
Average selected from the algorithm	1.45	4.5	4.77
Number of "AB" choices	0.2	1.77	2.68

Results

Choice

	random	fictitious	real
Average selected from the algorithm	1.45	4.5	4.77
Number of "AB" choices	0.2	1.77	2.68

Average time

- Manual choice – around 400s
- Algorithm (6 comparisons) – around 120s for real and 135 for fictitious

Results

Choice

	random	fictitious	real
Average selected from the algorithm	1.45	4.5	4.77
Number of "AB" choices	0.2	1.77	2.68

Average time

- Manual choice – around 400s
- Algorithm (6 comparisons) – around 120s for real and 135 for fictitious

Average gain (in %)

	manual	algorithm	final
Placebo		2.4	2.7
Real alternatives	2.7	3	2.8
Fictitious alternatives		3.1	3

Future efforts

1. Iterative approach

How to update the model when decision maker's preferences changed?

2. Machine learning

Huge amount of data to be used

1. Clustering
2. Patterns
3. Better questions