

# Meta učenje in AutoML: seminar

Ljupčo Todorovski

Univerza v Ljubljani, Fakulteta za upravo  
Institut Jožef Stefan, Odsek za tehnologije znanja (E8)

April 2019

# Pregled vsebine

## Nadgradnje meta modela za nadomestke

- Nastavitev parametrov meta modela
- Kombinacija z različnimi algoritmi

## Nadomestki in AutoML

- Uporaba meta modela za AutoML
- Primerjava meta modela z variantami SMBO

## Meta učenje

- AutoML in nadomestki
- Podatkovni tokovi
- Nevronske mreže

# Bledeči spomin pri učenju meta modela

Kako izbirati primere za učenje meta modela?

Trenutno stanje: Uporabniško določeni parametri

- $T_1, T_2$ : velikost množic za učenje modelov nadomestka in relevatorja
- $I_1, I_2$ : število novih primerov za ponovno učenje modelov

Odrta vprašanja in nadgradnje

- Opazovanje (bledeče) napake tako kot pri učenju iz podatkovnih tokov
- Vloga napake OOB (napak izven vreče)
- Izračun in upoštevanje zanesljivosti napovedi

# Izhodišča: učenje iz podatkovnih tokov

Posodobitveni operator: inkrementalen in dekrementalen

Kakšne posodobitve bi bile primerne za meta model?

Bledeča napaka s faktorjem bledenja  $\alpha$

$$E_f = \frac{\sum_{i=1}^n \alpha^{n-i} L(y_i, \hat{y}_i)}{\sum_{i=1}^n \alpha^{n-i}}$$

Metoda vrednotenja: zadržano ali napovedno zaporedno

Test za zaznavanje sprememb namesto  $l_1, l_2$

- Page-Hinkleyev test za napako
- Algoritem ADWIN za vrednosti ciljne funkcije

# Zanesljivost napovedi modelov

Kako izračunati in upoštevati zanesljivost napovedi?

## Trenutno stanje

Meta model ne upošteva zanesljivost napovedi.

## Odprta vprašanja

- Izračun zanesljivosti napovedi: ansambli
- Upoštevanje zanesljivosti pri izračunu relevantnosti

# Zanesljivost napovedi ansambla

Napoved naključnega gozda za primer  $\mathbf{x}$

$$\hat{y} = \frac{1}{B} \sum_{i=1}^B m_i(\mathbf{x})$$

- $B$  je število dreves v gozdu (ansamblu)
- $m_i$  so posamezna drevesa v gozdu
- $\hat{y}$  je pravzaprav povprečje  $\mu_{\hat{y}}$

Ocena zanesljivosti napovedi  $\sigma_{\hat{y}}$

$$\sigma_{\hat{y}}^2 = \frac{1}{B-1} \sum_{i=1}^B (m_i(\mathbf{x}) - \hat{y})^2$$

# Izhodišča: relevantnost

$$Relevance(x) = \begin{cases} \left(1 + \frac{F(x) - f_{min}}{f_{avg} - f_{min}}\right)^{-1} & ; F(x) \geq f_{min} \\ 1 & ; F(x) < f_{min} \end{cases}$$

- Formula pravzaprav računa oddaljenost od minimuma
- Relevantnost bi lahko (po analogiji SMBO) upoštevala  $\sigma_{\hat{y}}$

# Hevristike za določanje praga relevantnosti

Kako dinamično spreminjati vrednost praga relevantnosti?

$$D(x) = \begin{cases} 1 & ; \text{Relevance}(x) \geq T(h) \\ 0 & ; \text{Relevance}(x) < T(h) \end{cases}$$

Stopnja nadomeščanja  $RR$

$$RR = \frac{|\{(x_r, m_r, d_r) \in h : d_r = 0\}|}{|h|}$$

- $T$  dinamično prilagajamo tako, da vzdržujemo zaželeno stopnjo  $RR$
- Ustrezna formula prilagajanja?



# Izbira optimizacijskega algoritma

Kako robusten je meta model na izbiro optimizacijskega algoritma?

## Trenutno stanje

Algoritem za numerično optimizacijo *Differential Evolution*.

## Možnosti za nadaljevanje: bolj sodobni algoritmi

- CMA-ES: *Covariance Matrix Adaptation Evolution Strategy*, implementacija v Pythonu [github.com/CMA-ES/pycma](https://github.com/CMA-ES/pycma)
- Primerjava z ugnезdenim pristopom (Bajer in ost. 2018)
- Drugi algoritmi, glej spletni portal [coco.gforge.inria.fr](https://coco.gforge.inria.fr)

# Meta model in AutoML

Uporaba meta modela za optimizacijo parametrov učnih algoritmov?

## Odprta vprašanja

- Trenutno podprti le numerični parametri
- Povezano z izbiro optimizacijskega algoritma
- Empirična primerjava z variantami SMBO

# Algoritmi za numerično-celoštevilčno optimizacijo

Numerično-celoštevilčna (mešana) optimizacija  $m < k, m > 0$

$$\min_{x \in \mathbb{R}^{k-m} \times \mathbb{Z}^m} F(x)$$

Za  $m = 0$  običajna numerična optimizacija.

- Pomembno za iskanje optimalnih vrednosti diskretnih parametrov
- Problem: manjše število razpoložljivih algoritmov

Izhodišče za algoritme

- GAMBIT (Sadowski in ost. 2018)
- DASA (Korošec in ost. 2010)

# Primerjava meta modela z variantami SMBO

Primerjalna analiza splošnosti in zmogljivosti SMBO in meta modela?

Izhodišče za merjenje zmogljivosti

Benchmark problemi AutoML:

[www.automl.org/legacy/benchmarks.html](http://www.automl.org/legacy/benchmarks.html)

# Optimizacija vrednosti parametrov

Kdaj optimizirati vrednosti parametrov?

## Odprta vprašanja in možne teme

- Kdaj nastavljaati parametre za konkretno metodo, npr. naključni gozd
- Posebej za klasifikacijske in regresijske probleme
- Enako vprašanje za druge metode?
- Kako se meta modeli za različne metode razlikujejo?

## Izhodiščne raziskave

- (Ridd in Giraud-Carrier 2016)
- (Mantovani in ost. 2018)

# Primerjava dveh scenarijev

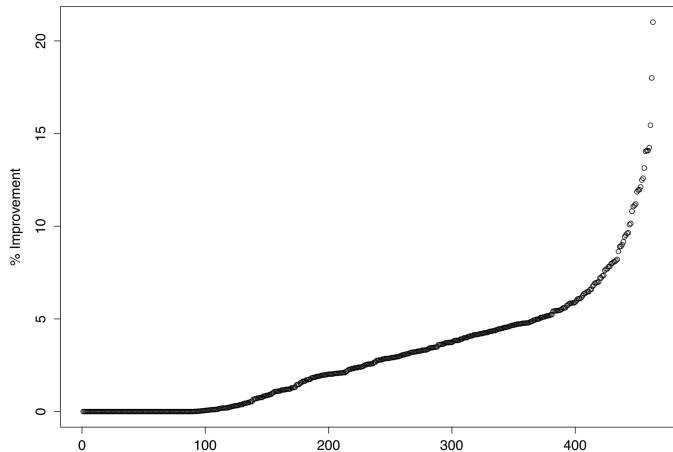
## Brez optimizacije parametrov

Najboljši algoritem s privzetimi vrednostmi parametrov.

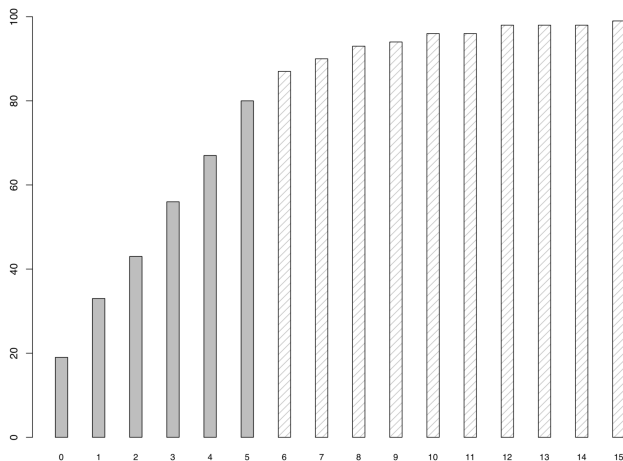
## Optimizacija parametrov

Najboljša konfiguracija dobljena z optimizacijo parametrov algoritmov.

# Izhodišča: Znatni učinki optimiziranja parametrov?



# Izhodišča: Neznatni učinki optimiziranja parametrov?





# Meta model: optimizirati ali ne?

```
nn_1 <= 0.902439
| joint_entropy <= 0.606044:  0 (14.0/2.0)
| joint_entropy > 0.606044
| | naive_bayes <= 0.952:  1 (208.0/30.0)
| | naive_bayes > 0.952:
| | | nn_1 <= 0.71134:
| | | | joint_entropy <= 2.08461:  0 (2.0)
| | | | joint_entropy > 2.08461:  1 (5.0)
| | | nn_1 > 0.71134:  0 (12.0/2.0)
nn_1 > 0.902439:  0 (85.0/15.0)
```

# Kombinacija meta učenja in autoML

Kako meta podatki lahko pomagajo pri optimizaciji vrednosti parametrov?

## Izhodišča

- Inicializacija optimizacije vrednosti parametrov (Wistuba in ost. 2015)
- Uporaba meta podatkov za gradnjo nadomestkov (Swersky in ost. 2013)

# Meta učenje za podatkovne tokove

## Splošen okvir za meta učenje iz podatkovnih tokov

- Meta spremenljivke za opis podatkovnih tokov?
- Kako upoštevati spreminjanje njihovih vrednosti skozi čas?
- Sprotna gradnja meta modela?

## Meta model za primerjavo delovanja različnih komponent

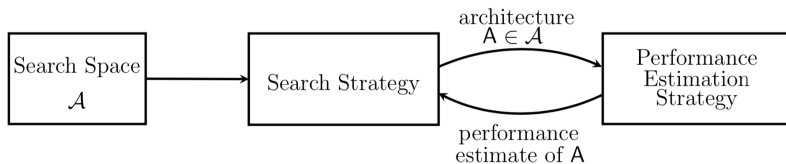
Primer: Kateri model ugotavljanja sprememb dela bolje kdaj?

# Meta učenje in nevronske mreže

Ali lahko nam meta učenje pomaga pri določanju topologije UNM?

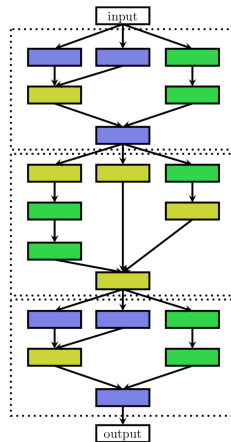
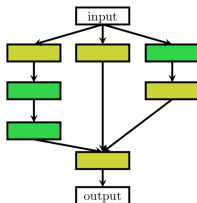
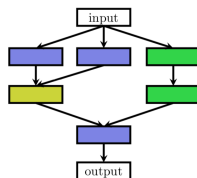
- (Elsken in ost. 2018): Pregled različnih metod

# Splošni okvir: iskanje optimalne topologije



# Strukturiranje preiskovalnega prostora topologij

Celice: pogosti topološki vzorci za gradnjo UNM



# Strategija vrednotenja zmogljivosti

Klasični pristopi (prečno preverjanje) preveč računsko zahtevni.

## Zato približno vrednotenje zmogljivosti

- Na majhnih vzorcih podatkovne množice: vzorčenje po primerih ali po spremenljivkah (pri slikah tudi po ločljivosti)
- Uporaba učnih krivulj: postopno povečevanje števila primerov
- Prenos (transfer) začetnih vrednosti uteži