

Optimalizace rozložení skladu pomocí evolučních algoritmů se surrogate modely

Feature Selection a Ensemble metody

David Zeman

ČVUT FEL

2026

Obsah

1 Úvod a motivace

2 Feature Selection

3 Ensemble modely

4 Experimenty a výsledky

5 Závěr

Problém MAPD (Multi-Agent Pickup and Delivery)

Problém:

- Sklad s mřížkovým rozložením (9×9)
- 49 pozic pro zboží, 4 agenti
- Agenti sbírají objednávky a doručují k výdejným místům
- Distribuce objednávek: **Zipf** (některé položky jsou populárnější)



Cíl:

- Najít optimální **rozložení zboží** ve skladu
- Maximalizovat **průchodnost** (throughput)
- Populární zboží blíže k okrajům \Rightarrow rychlejší obsluha

Červená = populární zboží

Problém prohledávacího prostoru

- Layout = permutace 49 položek $\Rightarrow 49! \approx 6 \times 10^{62}$ možností
- Jedna simulace trvá **2–3 sekundy**
- Evoluční algoritmus potřebuje **stovky až tisíce** evaluací

Řešení: Surrogate model

- Aproximace fitness funkcí pomocí **strojového učení**
- Extrakce **40 features** z layoutu (vzdálenosti, distribuce, kongesce, ...)
- Predikce throughputu **bez simulace** $\Rightarrow 1000\times$ rychlejší
- Hybridní přístup: občasné reálné evaluace pro aktualizaci modelu

Extrahované features (40 celkem)

Distance-based:

- Vážený průměr vzdálenosti k okraji
- Směrodatná odchylka vzdáleností
- Max vzdálenost top 10% položek
- Průměr top 5/20 položek

Position-based:

- Počet populárních položek na okraji
- Podíl v rozích
- Kvadrantové rozložení (NW, NE, SW, SE)

Distribution-based:

- Entropie vzdáleností
- Gini koeficient
- Pravděpodobnostní hmota na okraji

Congestion/clustering:

- Skóre kongesce
- Lokální hustota populárních položek
- Bottleneck skóre
- Path overlap

Důležitost features (Random Forest)

Top 10 nejdůležitějších features:

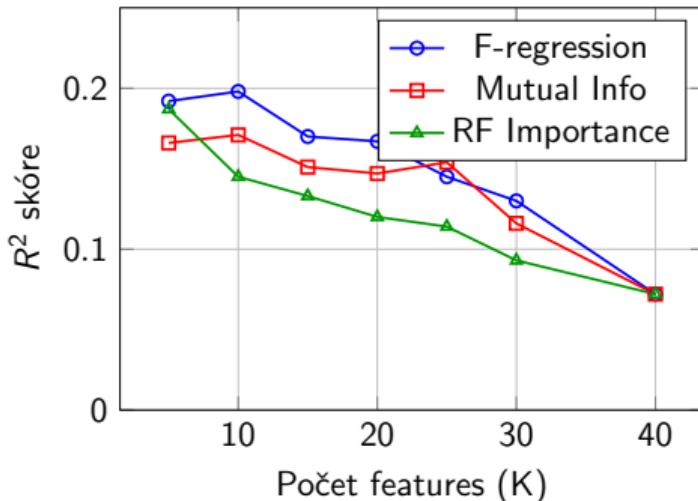
- ① edge_prob_mass (0.142)
- ② dist_gini (0.059)
- ③ quad_SE (0.050)
- ④ spread_x (0.050)
- ⑤ quad_NW (0.044)
- ⑥ hotspot (0.041)
- ⑦ dispersion (0.040)
- ⑧ skewness (0.038)
- ⑨ quadrant_balance (0.037)
- ⑩ quad_SW (0.037)

Klíčové zjištění

edge_prob_mass – pravděpodobnostní hmota na okraji – je **nejdůležitější prediktor throughputu**.

Populární položky blízko okraje ⇒ vyšší throughput

Vliv počtu features na kvalitu modelu



Výsledky

- Optimální počet features: $K = 10$
- R^2 se zvyšuje z 0.072 na **0.198**
- Příliš mnoho features \Rightarrow overfitting
- F-regression dosahuje nejlepších výsledků

Zlepšení

Redukce z 40 na 10 features:
+175% zlepšení R^2 skóre

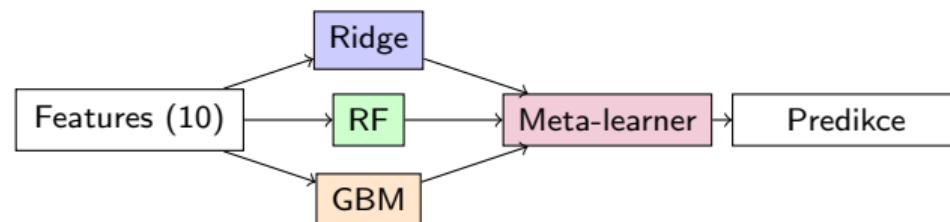
Ensemble metody

Voting Ensemble

- Kombinace 3 modelů:
 - Ridge Regression
 - Random Forest
 - Gradient Boosting
- Výstup = **průměr** predikcí
- Robustní, rychlý

Stacking Ensemble

- Stejné 3 základní modely
- **Meta-learner** (Ridge) kombinuje predikce
- Trénován pomocí cross-validation
- Lepší predikce, složitější trénink



Porovnání modelů – R^2 skóre

Model	Features	R^2
green!20 Voting Ensemble	10	0.166 ± 0.056
Baseline GP	40	0.152 ± 0.040
Ridge (selected)	10	0.145 ± 0.048
Baseline RF	40	0.145 ± 0.038
RF (selected)	10	0.136 ± 0.055
Stacking Ensemble	10	0.123 ± 0.019
GBM (selected)	10	0.084 ± 0.099
Baseline Ridge	40	0.072 ± 0.075
Baseline Linear	40	0.027 ± 0.105

Poznámka

Voting Ensemble s 10 features překonává všechny baseline modely se 40 features!

Nastavení experimentů

Parametry skladu:

- Velikost: 9×9
- Počet zboží: 49
- Počet agentů: 4
- Planner: A*
- Distribuce: Zipf 1.2

Parametry EA:

- Velikost populace: 20
- Počet generací: 50
- Reálná evaluace: každých 5 generací
- Počáteční vzorky: 300
- Simulační kroky: 300

Hybridní přístup

- ① Sběr 300 vzorků pro trénink surrogate modelu
- ② Evoluční algoritmus používá surrogate pro rychlé evaluace
- ③ Každých 5 generací: reálná evaluace pro aktualizaci modelu

Výsledky – Porovnání metod

Metoda	Fitness	Real evals	Čas
Random (best)	0.4800	–	–
Random (mean)	0.3399	–	–
Greedy	0.4400	–	–
EA bez surrogate	0.5500	1020	2608s
yellow!30 EA + Voting Ensemble	0.5600	168	451s
green!30 EA + Stacking Ensemble	0.5800	168	630s

Speedup

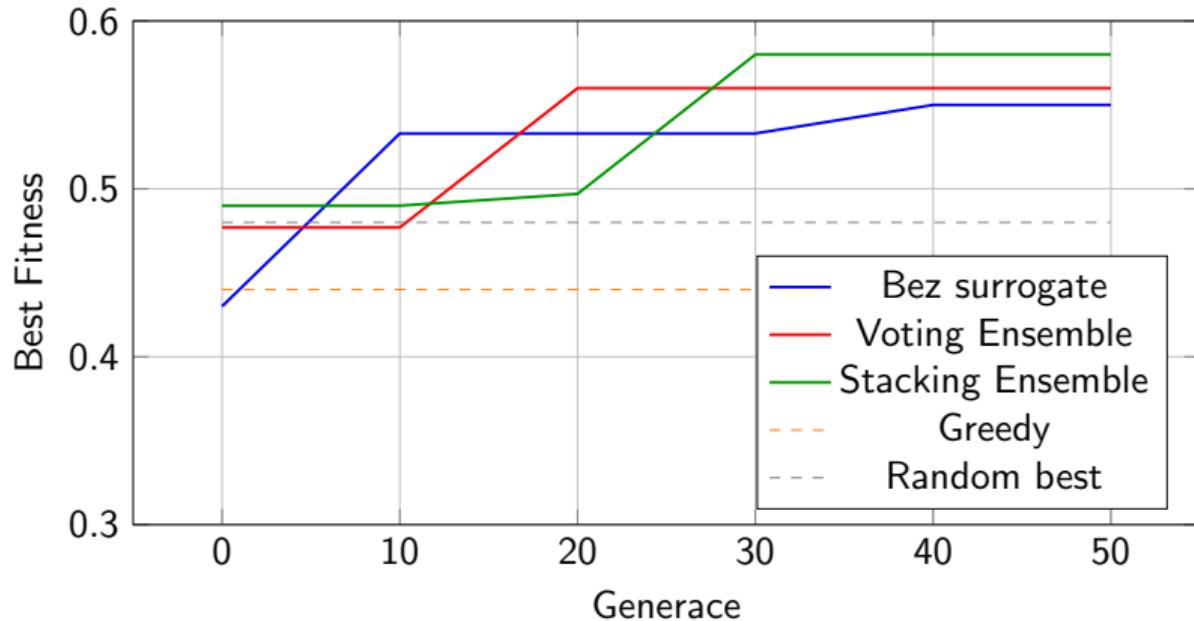
- Voting: **5.79×** rychlejší
- Stacking: **4.14×** rychlejší

Nejlepší metoda

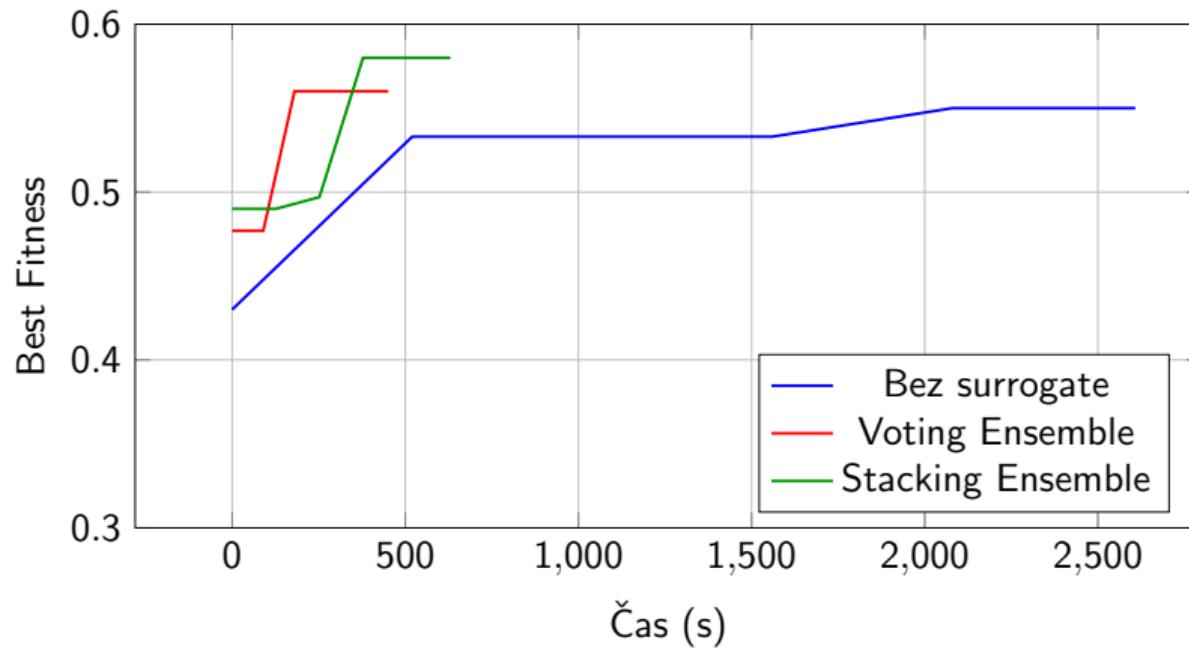
Stacking Ensemble

- Fitness: 0.5800 (+5.5%)
- 83.5% redukce evaluací

Konvergence EA



Fitness vs. výpočetní čas



Klíčový závěr: Surrogate metody dosahují lepších výsledků za zlomek času!

Shrnutí výsledků

Feature Selection

- Redukce 40 → 10 features
- Zlepšení R^2 : +175%
- Klíčová feature: edge_prob_mass

EA s surrogate

- Fitness: +5.5% vs bez surrogate
- Speedup: 4–6× rychlejší
- 83.5% méně reálných evaluací

Ensemble modely

- Voting: rychlý, robustní
- Stacking: nejlepší predikce
- Překonávají single modely

Hlavní přínos

Kombinace **feature selection** a **ensemble surrogate** umožňuje efektivní optimalizaci rozložení skladu s výraznou úsporou výpočetního času.

- **Adaptivní feature selection** – dynamický výběr features během evoluce
- **Transfer learning** – přenos znalostí mezi různými velikostmi skladů
- **Multi-objective optimalizace** – throughput vs. kolize vs. energetická náročnost
- **Online learning** – průběžná adaptace surrogate modelu
- **Větší sklady** – škálovatelnost na reálné warehouse rozměry

Praktické aplikace

- E-commerce fulfillment centra (Amazon, Alza, ...)
- Automatizované sklady s AGV roboty
- Optimalizace logistických operací

Děkuji za pozornost!

Otázky?

david.zeman@fel.cvut.cz