

Zadania diplomových prác 2025

Vedúci práce:

Samuel Rosa

M-240

<http://www.iam.fmph.uniba.sk/ospm/Rosa/teaching.htm>

samuel.rosa@fmph.uniba.sk

Dizajn experimentov: Na katedre sa viacerí (Harman, Filová, Somogyi, ja) venujeme optimálnemu dizajnu experimentov. Preto aj niektoré témy diplomoviek sú z tejto oblasti.

Dizajn (alebo navrhovanie) experimentov je oblasť štatistiky zaoberajúca sa otázkami typu: máme 10 políčok na sadenie kukurice, a chceme zistiť závislosť jej výnosnosti od množstva závlahy a typu hnojiva. Aké množstvá vody a typy hnojiva máme zvoliť na jednotlivých políčkach, aby sme sa o tých vzťahoch dozvedeli čo najviac? Technicky sa to premietne do čo najpresnejšieho odhadovania biet v lineárnej regresii, čomu zodpovedá minimalizácia nejakej funkcie variančnej matice MNŠ odhadu (napr. $\min \det((\mathbf{X}^T \mathbf{X})^{-1})$).

Zoradenie tém: témy sú zoradené približne v klesajúcej postupnosti môjho záujmu, takže skoršie čísla budem viesť podstatne radšej ako neskoršie.

1 Pažravé metódy v dizajne experimentov

Ľudskou rečou:

Dizajn experimentov: viď úvodný odsek.

Pažravé (greedy) metódy sú najjednoduchšie optimalizačné metódy na konštrukciu dizajnu. Konštruujú ho tak, že v každom kroku pridajú takú kombináciu nastavení (voda-hnojivo), ktorá najviac zníži príslušnú účelovú funkciu. Je to klasický postup (Sagnol 2013), ale v dizajne a oblastiach blízkyh dizajnu sa prednedávnom vyskytli rôzne zaujímavé modifikácie, napr.

- aproximácie rôznych výpočtov vnútri pažravých metód, aby boli rýchlejšie – využívajúc myšlienky ako v subsamplingu (Predari et al., 2023)
- skupinovo pažravé metódy a ich modifikácie, aby boli výsledky kvalitnejšie (Jiang et al. 2019, Nagata et al. 2023)
- nové dôkazy kvality (Chamon a Ribeiro, 2017)

Cieľom diplomovky by bolo (i) implementovať niektoré z týchto metód, s dôrazom na to, aby boli implementované efektívne, (ii) previesť teóriu o týchto metódach z oblastí blízkyh dizajnu do oblasti dizajnu experimentov. Téma umožňuje aj kreativitu, keďže pažravé metódy sa dajú modifikovať množstvom spôsobov podobných tým uvedeným.

Tento rok som z tejto témy viedol diplomovku (<https://opac.crzp.sk/?fn=detailBiblioFormChildQE789&sid=226A7A544DA3649F6F5139E9F6D0&seo=CRZP-detail-kniha>), ktorá však iba priamočiaro (a neefektívne) previedla jednu skupinovú pažravú metódu do sveta dizajnu experimentu.

Poznámka: Ide o serióznu tému dotýkajúcu sa súčasného výskumu. Samotná téma je ale pre dobrého študenta v pohode zvládnuteľná.

Požiadavky: Schopnosť študovať anglickú literatúru, slušné matematické zdatnosti a dobrá schopnosť programovať.

Literatúra:

- Sagnol, G. (2013). Approximation of a maximum-submodular-coverage problem involving spectral functions, with application to experimental designs. *Discrete Applied Mathematics*, 161(1-2), 258-276.
- Predari, M., Berner, L., Kooij, R., & Meyerhenke, H. (2023). Greedy optimization of resistance-based graph robustness with global and local edge insertions. *Social Network Analysis and Mining*, 13(1), 130.
- Jiang C., Chen Z., Su R., Soh Y. C.: Group Greedy Method for Sensor Placement . *IEEE Transactions on Signal Processing*, vol. 67, no. 9, pp. 2249-2262, 1 May, 2019.
- Nagata T., Yamada K., Nakai K., Saito Y., Nonomura T.: Randomized Group- Greedy Method for Large-Scale Sensor Selection Problems. *IEEE Sensors Journal*, vol. 23, no. 9, pp. 9536-9548, 1 May, 2023
- Chamon, L., & Ribeiro, A. (2017). Approximate supermodularity bounds for experimental design. *Advances in Neural Information Processing Systems*, 30.

2 ~~Multiplikatívny algoritmus v dizajne experimentov~~

Táto téma už je obsadená

Ľudskou rečou:

Dizajn experimentov: viď úvodný odsek.

Špeciálnou oblasťou dizajnu je tzv. aproximatívny dizajn, kde neurčujeme počty pokusov (napr. 2 pokusy pri jednej kombinácii voda-hnojivo, 1 pokus pri inej atď.), ale ľubovoľné reálne pomery (napr. 10% zo všetkých pokusov pri tejto kombinácii, 7.1478% pri druhej atď.). Výsledný optimalizačný problém má tú výhodu, že je spojitý a konvexný, a teda celkom dobre riešiteľný.

Jedným z klasických optimalizačných metód je multiplikatívny algoritmus (viď napr. Dette et al. 2008). Cieľom diplomovky je vylepšiť tento algoritmus. Možné vylepšenia sú napríklad:

- Niektoré výpočty vnútri počítať iba približne, aby sa algoritmus zrýchlil.

- Mierne upraviť metódu, napr. zahrnúť tzv. momentum (čo je istý štandardný trik v gradientných optimalizačných metódach, ktorý sa možno po vhodnej úprave dá použiť aj v tomto algoritme).
- V istej triede článkov autori, ktorí zrejme nepoznajú multiplikatívny algoritmus, začali používať niečo ako multiplikatívny algoritmus: “projected entropic mirror descent” (Allen-Zhu et al. 2021). Dá sa pozrieť sa na to, aký je vzťah týchto dvoch algoritmov (teda či sa len javia byť podobné, alebo aj ozaj sú, a aké sú ich praktické kvality) a či je možné použiť postupy/myšlienky z druhého algoritmu na vylepšenie alebo dokonca nahradenie multiplikatívneho.

Nie je cieľ zrealizovať všetky z týchto vylepšení (kvalitne vypracované jedno z nich je postačujúce, pokiaľ sa ukáže byť ako-tak zmysluplným). Samozrejmosťou súčasťou je efektívne naprogramovanie algoritmu (ktorý je ale veľmi priamočiary) a numerické porovnanie originálnej verzie s vylepšeniami alebo s inými verziami.

Poznámka: Ide o serióznú tému dotýkajúcu sa súčasného výskumu. Samotná téma je ale pre dobrého študenta v pohode zvládnuteľná.

Požiadavky: Schopnosť študovať anglickú literatúru, slušné matematické zdatnosti a schopnosť programovať.

Literatúra:

- Dette, H., Pepelyshev, A., & Zhigljavsky, A. (2008). Improving updating rules in multiplicative algorithms for computing D-optimal designs. *Computational Statistics & Data Analysis*, 53(2), 312-320.
- Allen-Zhu, Z., Li, Y., Singh, A., & Wang, Y. (2021). Near-optimal discrete optimization for experimental design: A regret minimization approach. *Mathematical Programming*, 186, 439-478.

3 Metódy hlavných komponentov

Anglicky: Principal component analyses

Anotácia: Metóda hlavných komponentov je základná metóda menejrozmernej reprezentácie mnohorozmerných dát, používaná napríklad na vykreslenie takých dát do dvojrozmerného obrázku. Hoci je užitočná a široko používaná, má rôzne nedostatky, medzi ktoré patrí napríklad nerobustnosť vzhľadom na outliery a niekedy ťažšia interpretovateľnosť. Preto boli vyvinuté jej upravené verzie, ako napríklad robustná metóda hlavných komponentov a riedka (sparse) metóda hlavných komponentov, ktoré tieto nedostatky riešia. Cieľom práce je naštudovať tieto upravené metódy hlavných komponentov a preskúmať ich správanie na reálnych alebo simulovaných dátach.

Požiadavky: Schopnosť študovať anglickú literatúru.

4 Kvantifikácia vplyvov premenných

Anglicky: Feature importance quantification

Anotácia: Narozdiel od klasických metód, moderné metódy strojového učenia typicky poskytujú kvalitné

predikcie za cenu práce s ťažko interpretovateľnými modelmi (čiernymi skrinkami). Preto boli vyvinuté rôzne metódy, ktoré sa snažia spätne interpretovať získané modely: najmä tým, že kvantifikujú mieru vplyvu jednotlivých premenných na ten-ktorý model. Medzi tieto metódy patria napríklad SHAP (Shapley Additive Explanation), LIME (Local Interpretable Model-agnostic Explanations) a PDP (Partial Dependence Plots). Cieľom práce je naštudovať niektoré metódy slúžiace na kvantifikáciu vplyvov premenných, čitateľsky prístupne ich spísať do uceleného textu a preskúmať ich správanie na reálnych, prípadne simulovaných, dátach.

Požiadavky: Schopnosť študovať anglickú literatúru.