

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 872

VARIJACIJSKO UČENJE NA ZAŠUMLJENIM OZNAKAMA

Dominik Jambrović

Zagreb, lipanj, 2025.

Zagreb, 3. ožujka 2025.

DIPLOMSKI ZADATAK br. 872

Pristupnik: **Dominik Jambrović (0036534818)**

Studij: Računarstvo

Profil: Računarska znanost

Mentor: prof. dr. sc. Siniša Šegvić

Zadatak: **Varijacijsko učenje na zašumljenim oznakama**

Opis zadatka:

Raspoznavanje slika važan je problem računalnog vida s mnogim zanimljivim primjenama. U posljednje vrijeme stanje tehnike postižu duboki modeli zasnovani na konvolucijama i slojevima pažnje. Međutim, standardni postupci teško se nose sa zašumljenim oznakama. U okviru rada, potrebno je odabrati okvir za automatsku diferencijaciju te upoznati biblioteke za rukovanje tenzorima i slikama. Proučiti i ukratko opisati postojeće duboke arhitekture za raspoznavanje slika s posebnim naglaskom na prednaučene samonadzirane modele. Odabrati slobodno dostupne skupove slika te oblikovati podskupove za učenje, validaciju i testiranje. Formulirati optimizacijski cilj s latentnim predikcijama čistih razreda te predložiti rješenje utemeljeno na varijacijskoj aproksimaciji te maksimiziranju očekivanja. Komentirati učinkovitost učenja i zaključivanja. Predložiti pravce za budući rad. Radu priložiti izvorni i izvršni kod razvijenih postupaka, ispitne slijedove i rezultate, uz potrebna objašnjenja i dokumentaciju. Citirati korištenu literaturu i navesti dobivenu pomoć.

Rok za predaju rada: 4. srpnja 2025.

Zahvale!

Sadržaj

1. Uvod	2
2. Problem zatrovanih podataka	4
3. Problem zašumljenih podataka	5
4. Samonadzirano učenje	6
5. Algoritam maksimizacije očekivanja	7
6. Transportni problem	8
7. VIBE	9
8. Skup podataka	10
9. Eksperimenti	11
10. Zaključak	12
Literatura	13
Sažetak	15

1. Uvod

Duboki modeli koriste se u brojnim aspektima naše svakodnevice. Pri razvoju i učenju modela, pažnju prije svega posvećujemo performansama na neviđenim podacima - želimo naučiti modele koji dobro generaliziraju. Drugim riječima, želimo da modeli daju ispravna predviđanja za viđene, ali i za neviđene podatke. Ovime osiguravamo da naša rješenja imaju primjenu i van laboratorijskih uvjeta u kojima se uče.

U procesu razvoja modela za određeni zadatak strojnog učenja, osim odabira arhitekture, algoritma učenja i hiperparametara, veliku ulogu igraju podatci na kojima učimo. Općenito govoreći, prikupljanje i označavanje podataka jedan je od najskupljih dijelova procesa razvoja rješenja za nekih problem. Važno je da prikupljeni podatci što realističnije predstavljaju stvarne situacije s kojima će se naš model susretati tj. da distribucija podataka odgovara stvarnoj distribuciji situacija koje prikazuju. Dodatno, pokazuje se da duboki modeli uz dovoljan kapacitet mogu naučiti ispravno predviđati oznake čak i za nasumično označene podatke [1], tako da je veoma važno da su prikupljeni podatci što točnije označeni.

Područje računalnog vida [2] bavi se razvojem algoritama i modela za brojne zadatke raspoznavanja i razumijevanja slika. Najčešći zadatak je klasifikacija slika - model na ulazu dobiva sliku, a na izlazu treba predvidjeti razred koji odgovara ulaznom primjeru. Iako postoje brojni skupovi slikovnih podataka koji se mogu koristiti za učenje i evaluaciju modela, za konkretne zadatke u većini slučajeva trebamo prikupiti i označiti vlastite slike. Pritom postoji nekoliko čestih opasnosti: prisutnost zatrovanih [3] ili zašumljenih [4] podataka.

Kada govorimo o trovanju podataka, maliciozni agent u skup podataka dodaje zatrovane podatke s ciljem manipulacije izlaza naučenog modela za određene ulaze. S druge strane, anotator podataka bez zlih namjera određenim podacima može pridijeliti netočne oznake, time dodajući zašumljene podatke u skup. Kroz vrijeme, razvili su se brojni algoritmi za obranu modela od zatrovanih [5, 6, 7] odnosno zašumljenih [8, 9] podataka. Ipak, većina radova se fokusira na samo jedan od ovih problema, a ne na razvoj algoritma koji se može nositi s oba problema.

Cilj ovog rada je reproducirati i poboljšati rezultate okvira za obranu od zatrovanih podataka imena VIBE [10]. Osim ovoga, cilj je i primijeniti VIBE na problem zašumljenih podataka. Pritom VIBE evaluiramo na nekoliko čestih vrsta trovanja odnosno zašumljivanja podataka kako bi se osigurala robusnost okvira. Dodatno, cilj je usporediti VIBE sa stanjem tehnike (engl. *state of the art* - *SotA*) za problem zašumljenih podataka.

2. Problem zatrovanih podataka

3. Problem zašumljenih podataka

4. Samonadzirano učenje

5. Algoritam maksimizacije očekivanja

6. Transportni problem

7. VIBE

8. Skup podataka

9. Eksperimenti

10. Zaključak

Literatura

- [1] C. Zhang, S. Bengio, M. Hardt, B. Recht, i O. Vinyals, “Understanding deep learning requires rethinking generalization”, *arXiv preprint arXiv:1611.03530*, 2016.
- [2] A. Voulodimos, N. Doulamis, A. Doulamis, i E. Protopapadakis, “Deep learning for computer vision: A brief review”, *Computational intelligence and neuroscience*, sv. 2018, br. 1, str. 7068349, 2018.
- [3] B. Biggio, B. Nelson, i P. Laskov, “Poisoning attacks against support vector machines”, *arXiv preprint arXiv:1206.6389*, 2012.
- [4] S. Gupta i A. Gupta, “Dealing with noise problem in machine learning data-sets: A systematic review”, *Procedia Computer Science*, sv. 161, str. 466–474, 2019.
- [5] Y. Li, X. Lyu, N. Koren, L. Lyu, B. Li, i X. Ma, “Anti-backdoor learning: Training clean models on poisoned data”, *Advances in Neural Information Processing Systems*, sv. 34, str. 14 900–14 912, 2021.
- [6] K. Huang, Y. Li, B. Wu, Z. Qin, i K. Ren, “Backdoor defense via decoupling the training process”, *arXiv preprint arXiv:2202.03423*, 2022.
- [7] Z. Zhang, Q. Liu, Z. Wang, Z. Lu, i Q. Hu, “Backdoor defense via deconfounded representation learning”, u *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023., str. 12 228–12 238.
- [8] S. Liu, Z. Zhu, Q. Qu, i C. You, “Robust training under label noise by over-parameterization”, u *International Conference on Machine Learning*. PMLR, 2022., str. 14 153–14 172.

- [9] H. Chen, A. Shah, J. Wang, R. Tao, Y. Wang, X. Li, X. Xie, M. Sugiyama, R. Singh, i B. Raj, “Imprecise label learning: A unified framework for learning with various imprecise label configurations”, *Advances in Neural Information Processing Systems*, sv. 37, str. 59 621–59 654, 2024.
- [10] I. Sabolić, M. Grcić, i S. Šegvić, “Seal your backdoor with variational defense”, *arXiv preprint arXiv:2503.08829*, 2025.

Sažetak

Varijacijsko učenje na zašumljenim oznakama

Dominik Jambrović

Sažetak...

Ključne riječi: prva ključna riječ; druga ključna riječ; treća ključna riječ