

# █ Pozitivno in neoznačeno učenje z generativnimi nasprotniškimi mrežami

Aleš Papič, Igor Kononenko, Zoran Bosnić

Univerza v Ljubljani, Fakulteta za računalništvo in informatiko, Večna pot 113, 1000 Ljubljana, Slovenija

{ales.papic, igor.kononenko, zoran.bosnic}@fri.uni-lj.si

## Izvleček

Količina ustvarjenih podatkov otežuje njihovo obdelavo. V primeru nadzorovanega učenja lahko označevanje učnih primerov predstavlja dolgotrajno in drago nalogo. V dvorazrednih problemih primere označimo kot pozitivne ali negativne. V tem delu predpostavljamo, da imamo na voljo majhno število pozitivnih primerov in večje število neoznačenih primerov. Cilj generativnega pozitivnega in neoznačenega učenja je ustvariti označene primere, kar predstavlja strategijo za reševanje tega problema. Kljub temu generativni pristopi prinašajo pomanjkljivosti, kot so visoka računska zahtevnost, nestabilno učenje in nezmožnost ustvarjanja popolnoma označenih podatkovnih množic. V prispevku predlagamo nov generativni pristop, ki temelji na pomožnih klasifikacijskih generativnih nasprotniških mrežah. Nenegativno pozitivno in neoznačeno tveganje integriramo kot pomožno funkcijo izgube, da se naučimo porazdelitve pozitivnih in negativnih primerov. Na priznanem naboru podatkov za pozitivno in neoznačeno učenje prikažemo najsodobnejše rezultate. Rezultati kažejo, da naš pristop dosega primerljivo točnost z obstoječimi pristopi, kljub preprostejši arhitekturi, ima visoko učno stabilnost in generira nabor označenih podatkov.

**Ključne besede:** pozitivno in neoznačeno učenje, delno nadzorovano učenje, generativne nasprotniške mreže, globoko učenje

## Abstract

The quantity of the data generated makes them difficult to process. In the case of supervised learning, labelling training examples may represent an especially tedious and costly task. In binary classification problems, examples are labelled as positive or negative. In this work, we assume that we have at our disposal a small number of positive and a larger number of unlabelled examples. Generative positive and unlabelled learning aims to generate labelled data, which constitutes a strategy to address this issue. Nonetheless, the generative approaches bring shortcomings, such as high computational cost, training instability and the inability to generate fully labelled datasets. We propose a novel generative approach based on an auxiliary classifier generative adversarial network. We integrate non-negative positive and unlabelled risk as an auxiliary loss to learn the distribution of positive and negative examples. We demonstrate the state-of-the-art performance on a common positive and unlabelled learning benchmark dataset. The results show that our approach achieves comparable performance as existing approaches despite its simple architecture, has high training stability and generates fully labelled data.

**Keywords:** Positive and unlabelled learning, partially supervised learning, generative adversarial networks, deep learning

## 1 UVOD

Naloge nadzorovanega učenja zahtevajo, da so učni primeri označeni, *npr.* kot pripadniki pozitivnega ali negativnega razreda, pred učenjem modela. Vsakodnevno pridobivanje označenih podatkov ni vedno samoumevno zaradi pomanjkanja človeških strokovnjakov, s tem povezanih stroškov, časovnih omejitev ali celo omejitev postopka zajemanja podatkov. Zaradi omenjenih praktičnih pomanjkljivosti je postalo podpodročje delno nadzorovanega učenja, pozitivno

in neoznačeno (*angl.* positive and unlabeled, PU) učenje, vse bolj proučevano.

PU učenje je binaren, pozitiven in negativen klasifikacijski problem, kjer so na voljo le označeni pozitivni in neoznačeni primeri. Elkan in Noto [6] sta neoznačene podatke obravnavala kot utežene pozitivne in negativne hkrati, kar je pozneje spodbudilo razvoj nepristranskih tehnik [4, 3, 8]. S tem predobdelava neoznačenih podatkov ni potrebna, kar poenostavi učni proces. Ne glede na to z omejenim

Tabela 1: Primerjava prednosti predstavljenih najsodobnejših PU pristopov. Oznaka (✓) poudarja, da metoda izpolnjuje dano merilo na levi [11].

Metoda	GenPU	PGAN	D-GAN
Predznanje ni potrebno		✓	✓
Primeren za zapletene nabore podatkov		✓	✓
Ustvari ustrezne pozitivne primere	✓		
Ustvari ustrezne negativne primere	✓		✓
Stabilnost učenja z uporabo SGD		✓	✓
Izvirna arhitektura GAN		✓	✓

naborom označenih podatkov obstaja nevarnost prevelikega prileganja, kadar se uporabljajo prilagodljivi modeli, kot so globoke nevronske mreže. Zato se v zadnjem času raziskave osredotočajo na uporabo generativnih nasprotniških mrež (*angl.* generative adversarial networks, GAN) zaradi rasti razpoložljivih neoznačenih podatkov [7, 2, 1]. GANi bogatijo učne primere z umetnimi, da je na voljo večji nabor označenih primerov, ki jih lahko uporabimo za nadzorovano učenje. Kljub temu imajo generativni pristopi PU učenja veliko računsko zahtevnost oziroma generirajo izključno negativne učne primere. To pa lahko predstavlja težavo med učenjem klasifikatorja v primerih, ko primanjkuje označenih pozitivnih primerov.

V ta namen predlagamo novo generativno PU ogrodje za generiranje označenih učnih primerov. Ogrodje uporablja pogojno latentno predstavitev prostora in je opremljeno s pomožno ocenitveno funkcijo za učenje iz PU podatkov. Poleg tega je predlagani pristop mogoče uporabiti na različnih znanih arhitekturah GAN. S poskusi na resničnih podatkih pokažemo, da naš pristop uspešno prepozna porazdelitev tako pozitivnih kot negativnih primerov.

## 2 SORODNA DELA

Du Plessis in sod. [3] so definirali konveksno PU učenje (uPU), toda Kiryo in sod. [8] so kasneje pokazali, da se prilagodljivi modeli, kot so nevronske mreže, pretirano prilegajo podatkom. To se zgodi, ko vrednost funkcije izgube postane negativna, zato so predlagali nenegativno različico (nnPU). Vendar pa uPU in nnPU zahtevata predhodno poznavanje po-

razdelitve podatkov. Chiaroni in sod. [1] so poudarili, da se porazdelitev primerov razlikuje med paketi (*angl.* batch), zaradi česar lahko stohastične tehnike optimizacije postanejo nestabilne, zlasti pri majhnih paketih. Zato se nedavne raziskave osredotočajo na generativne nasprotniške mreže (GANe), da bi podatke obogatili z umetnimi učnimi primeri in tako povečali označen učni nabor.

Mirza in sod. [10] so predlagali strategijo pogojnega GANa (CGAN), ki generatorju in diskriminatorju poda oznako razreda ter tvori z razredom pogojene umetne primere. Odena in sod. [11] so predlagali drugo različico, imenovano *pomožni klasifikator* GAN (AC-GAN), kjer ima diskriminator pomožno nalogo, da napove pogojno latentno informacijo. Skozi nasprotniški proces učenja se generator nauči skupne latentne predstavitve. Kljub temu oba pristopa potrebujeta označen nabor podatkov, česar pri PU učenju nimamo in predstavlja njuno glavno pomanjkljivost.

Velike količine neoznačenih podatkov otežujejo učni proces obstoječih PU metod. Zato so Hou in sod. [7] predlagali prvo generativno PU ogrodje (GenPU), ki vključuje vrsto generatorjev in diskriminatorjev, zaradi česar je računsko zahteven. Chiaroni in sod. [2] so predlagali Pozitiven-GAN (PGAN), ki temelji na originalni arhitekturi GAN. Zaradi pomanjkanja nadzora pa se ne more naučiti resnične negativne porazdelitve podatkov, kar so kasneje naslovili z metodo Divergentni-GAN (D-GAN) [1]. Ne glede na to je še vedno omejen na generiranje izključno negativnih primerov. To pa lahko predstavlja težavo pri učenju končnega klasifikatorja, ko je na voljo majhen nabor označenih primerov. Tabela 1 povzema ključne lastnosti generativnih PU pristopov.

## 3 POGOJNO GENERATIVNO PU UČENJE

Naše delo naslavlja omejitvi ogrodja GenPU, visoko računsko zahtevnost in nestabilnost učenja, ki sta posledici zapletene arhitekture. Izvirno GAN arhitekturo dopolnimo z dodatnim pomožnim klasifikatorjem v novem pogojnem generativnem PU ogrodju. Le-ta je zmožen generiranja primerov obeh razredov, kar pa ne velja za obstoječa PGAN in D-GAN. Omenjena lastnost je potrebna za probleme, pri katerih nam primanjkuje označenih pozitivnih primerov.

Naš pristop je zgrajen na obstoječem AC-GANu, ki lahko generira označen nabor podatkov. Pomožni klasifikator optimizira log-verjetje pravega razreda  $L_c$  [11], zato potrebuje označene primere iz

vsakega razreda. Da pa bi se lahko učili iz PU nabora, uvedemo nenegativno PU tveganje ( $L_{nnPU}$ ) [8] kot dodatno pomožno funkcijo izgube.  $L_{nnPU}$  omogoča oceno izgube za neznane negativne primere, ki se nahajajo v vhodnih podatkih. Funkcijo izgube pomožnega klasifikatorja tako sestavlja utežena vsota  $\lambda_1 L_C + \lambda_2 L_{nnPU}$  kjer je  $\lambda_1 = 1 - \lambda_2$ .  $L_C$  merimo na umetnih primerih, saj so njihove oznake na voljo v latentni predstavitvi, medtem ko  $L_{nnPU}$  merimo izključno na učni množici. Pri empiričnem ovrednotenju modela je parameter  $\lambda_1 = 0.5$ .

Cilj pomožnega klasifikatorja je usmeriti generator k pravilni uporabi podane pogojne informacije. Ne glede na to moramo vedeti, kakšna je porazdelitev neoznačenih podatkov, kar lahko ocenimo neposredno iz podatkov samih [5]. Ogrodje nam omogoča, da generiramo popolnoma označen nabor pozitivnih in negativnih primerov, iz pozitivnih in neoznačenih. Predlagan pristop je dvostopenjski, kar pomeni, da se najprej naučimo generirati umetne primere, le-te pa uporabimo za učenje poljubnega obstoječega binarnega klasifikatorja. Ob tem ima naš pristop manj računskih omejitev in večjo stabilnost.

#### 4 REZULTATI IN RAZPRAVA

Učinkovitost generativnih pristopov pokažemo z empiričnim ovrednotenjem na podatkovni množici MNIST [9]. Med vrednotenjem se domneva, da je znana porazdelitev podatkov ( $\pi_p$ ). Izbrali smo dve številki s po 5000 slikami, ki smo jih uporabili za izdelavo nabora podatkov PU. Da bi bil problem bolj zahteven, smo izbrali številki 3 in 5 zaradi njune vizualne podobnosti. Izvedli smo štiri ponovitve, vsako z različnim številom označenih primerov, in oceni-

li uspešnost binarnega klasifikatorja, naučenega na umetnih primerih s 3-kratnim prečnim preverjanjem (glej Tabela 2). Za oceno učinkovitosti je bil uporabljen binarni klasifikator z dvema skritima plastema, vsaka z 256 nevroni. Aktivacijska funkcija za vsako plastjo je ReLU, razen zadnje, kjer je uporabljen sigmoid. Splošna arhitektura GAN in učni parametri so bili vzeti iz literature [7].

Rezultati kažejo, da klasifikator, naučen na podatkih predlaganega pristopa, dosega primerljivo točnost kot tisti, ki je bil naučen na podatkih PGANA in D-GANA. Na naše presenečenje, ko je na voljo le deset primerov z oznako, PGAN deluje najbolje. Hou in sod. [7] so ugotovili, da se nnPU začne pretirano prilegati, ob manjšem številu označenih primerov, zato se obnaša nestabilno. V našem ogrodju se je pojavila enaka nestabilnost, kar je vplivalo na generator, ki mu ni uspelo pravilno uporabiti podane pogojne informacije.

Že s 50 označenimi primeri pa je naš pristop presegel obstoječe. GenPU se je izkazal najslabše zaradi nestabilnega učenja in posledično nezmožnosti se naučiti porazdelitve negativnih primerov. Slika 1 prikazuje generirane številke s predlaganim pristopom.



Slika 1: Vizualizacija umetnih pozitivnih in umetnih negativnih števk, ustvarjenih s predlaganim pristopom.

Tabela 2: Klasifikacijska točnost klasifikatorja na nalogi razlikovanja števk, 3 in 5 ob različnem številu označenih podatkov.

PODATKI			METODA			
$N_p$	$N_o$	$\pi_p$	PGAN	D-GAN	GenPU	Pog. gen. PU učenje (naš)
1000	9000	0.444	<b>0.93 ± 0.01</b>	<b>0.93 ± 0.01</b>	0.47 ± 0.00	0.92 ± 0.00
100	9900	0.495	0.80 ± 0.06	0.82 ± 0.04	0.47 ± 0.00	<b>0.86 ± 0.04</b>
50	9950	0.497	0.72 ± 0.08	0.65 ± 0.13	0.47 ± 0.00	<b>0.74 ± 0.03</b>
10	9990	0.499	<b>0.69 ± 0.03</b>	0.61 ± 0.07	0.46 ± 0.00	0.59 ± 0.05

## 5 ZAKLJUČEK

Količina ustvarjenih podatkov se dnevno povečuje, kar otežuje njihovo obdelavo. Označevanje učnih primerov predstavlja še posebej dolgotrajno in drago, včasih tudi nemogočo nalogo. V članku predstavimo novo generativno ogrodje, ki zmanjšuje računsko zahtevnost in posledično povečuje učno stabilnost. Ob tem je primeren za širok nabor obstoječih arhitektur GAN in daje označen nabor podatkov. Rezultati so pokazali, da obstoječe metode največkrat dosežejo nižjo točnost kot predlagan pristop, še posebej, ko je število označenih primerov omejeno. Vendar pa je naš pristop pokazal težave pri zelo majhnem številu označenih primerov. To je posledica čezmernega prileganja pomožnega klasifikatorja, kar bomo obravnavali v prihodnje. Ne glede na to je predlagani pristop s sestavljeno kriterijsko funkcijo preprostejši, saj nadomešča kompleksno arhitekturo ogrodja GenPU.

Predlagan pristop je primeren za številne probleme v medicini, *npr.* pri odkrivanju redkih bolezni, v priporočilnih sistemih za prepoznavanje zavajajočih ocen in v bančništvu za zaznavanje rizičnih posojil. Pomaga lahko tudi pri anonimizaciji občutljivih podatkov, kjer je mogoče nadomestiti vhodne podatke z umetnimi pri tem pa jih loči v za nas bolj (pozitivno) in manj (negativno) zanimivo skupino.

V prihodnje nameravamo pristop uporabiti za odkrivanje nenapovedanih obrokov pri bolnikih s sladkorno boleznijo tipa 1, saj je problem primeren za PU učenje. Sedanja umetna trebušna slinavka nima te zmožnosti, a bi obstoj takšnega sistema zelo pripomogel bolnikom.

## ZAHVALA

To delo podpira Javna agencija za raziskovalno dejavnost Republike Slovenije (ARRS), sredstva za mlade raziskovalce (53630).

## LITERATURA

- [1] Florent Chiaroni, Ghazaleh Khodabandelou, Mohamed-Cherif Rahal, Nicolas Hueber, and Frédéric Dufaux. Generating relevant counter-examples from a positive unlabeled dataset for image classification. *arXiv preprint arXiv:1910.01968*, 2019.
- [2] Florent Chiaroni, Mohamed-Cherif Rahal, Nicolas Hueber, and Frédéric Dufaux. Learning with a generative adversarial network from a positive unlabeled dataset for image classification. In *IEEE ICIP*, pages 1368–1372. IEEE, 2018.
- [3] Marthinus Du Plessis, Gang Niu, and Masashi Sugiyama. Convex formulation for learning from positive and unlabeled data. In *ICML*, pages 1386–1394, 2015.
- [4] Marthinus C Du Plessis, Gang Niu, and Masashi Sugiyama. Analysis of learning from positive and unlabeled data. In *NIPS*, pages 703–711, 2014.
- [5] Marthinus Christoffel Du Plessis and Masashi Sugiyama. Class prior estimation from positive and unlabeled data. *IEICE*, 97(5):1358–1362, 2014.
- [6] Charles Elkan and Keith Noto. Learning classifiers from only positive and unlabeled data. In *KDD'08*, pages 213–220, 2008.
- [7] Ming Hou, Brahim Chaib-Draa, Chao Li, and Qibin Zhao. Generative adversarial positive-unlabeled learning. In *IJCAI*, pages 2255–2261. AAAI Press, 2018.
- [8] Ryuichi Kiryo, Gang Niu, Marthinus C Du Plessis, and Masashi Sugiyama. Positive-unlabeled learning with non-negative risk estimator. In *NIPS*, pages 1675–1685, 2017.
- [9] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- [10] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.
- [11] Augustus Odena, Christopher Olah, and Jonathon Shlens. Conditional image synthesis with auxiliary classifier gans. In *ICML*, pages 2642–2651. Proceedings of Machine Learning Research, 2017.

▪

**Aleš Papič** je mladi raziskovalec in doktorski študent na Fakulteti za računalništvo in informatiko, Univerze v Ljubljani. Njegovo raziskovalno področje vključuje strojno učenje, globoko učenje in generativne nasprotniške mreže. Prav tako je asistent pri predmetu Osnove umetne inteligence.

▪

**Igor Kononenko** je doktor računalniških znanosti in redni profesor na Fakulteti za računalništvo in informatiko Univerze v Ljubljani ter predstojnik Laboratorija za kognitivno modeliranje. Njegova raziskovalna področja so umetna inteligenca, strojno učenje, nevronske mreže in kognitivno modeliranje. Je (so)avtor 225 člankov na teh področjih ter 13 učbenikov (dve knjigi izšli v Angliji).

▪

**Zoran Bosnić** je profesor na Fakulteti za računalništvo in informatiko Univerze v Ljubljani. Raziskovalno se ukvarja z umetno inteligenco, zlasti s strojnimi učenjem. Osredotoča se pretežno na učenje iz podatkovnih tokov in na interdisciplinarne aplikacije strojnega učenja. Na tem področju je tudi (so)avtor okoli 70 znanstvenih člankov.