

RAPORTARE ȘTIINȚIFICĂ

Învățare prin transfer de domeniu pentru estimarea valorii estetice a unei imagini

CORNELIU FLOREA - DIRECTOR PROIECT
Universitatea Politehnica București, Laboratorul de Analiza și Prelucrarea Imaginilor
21 noiembrie 2020

1 Acțiuni planificate

Conform planului de realizare a proiectului această etapă conține activitatea: "Algoritmi pentru învățare prin transfer și pentru învățare semi-supervizată ce presupune: (I) Identificare rezultate relevante apărute de la depunerea proiectului. (II) Identificare baze de date potrivite. (III) Inițializare construcție de transfer de cunoștințe prin antrenare duală (predicție și grupare). (IV) Raportare și diseminare. În continuare, în acest raport sunt rezumate eforturile noastre în atingerea obiectivelor asumate în proiectul de față. Abordarea urmărește planul propus și un sistem bazat pe rețele convoluționale adânci antrenate dual (predicție și grupare) pe baza principiului *marjei largi* asumat în propunerea de proiect.

2 Soluția investigată

Am investigat o metodologie pentru a antrena o rețea convoluțională adâncă într-o manieră complet supervizată pentru probleme de clasificare mutual exclusive. Urmărim să creștem cantitatea de date cu unele sintetice având etichete soft (probabilistice). Acest lucru se realizează luând în considerare o grupare (clustering) controlată a descriptorilor. În consecință, rețeaua antrenată include un strat care acționează ca descriptor discriminativ. Pentru implementare am folosit un ResNet-18 [2], iar descriptorii sunt extrași din ultimul strat înainte de cel de decizie, după aplatizare (flattening).

Pentru probleme de clasificare, eticheta inițială este dată ca scalar (valoare categorială) și ulterior considerată în format vectorial \mathbf{y}_i (*one-hot encoding*). Ne propunem să calculăm probabilitățile de apartenență (soft) la clasă care sunt notate cu $\mathbf{y}_i = [y_i^1, y_i^2, \dots, y_i^C]$.

Grupare supervizată. Având o intrare \mathbf{x}_i , descriptorul asociat \mathbf{e}_i și predicția y_i , sistemul antrenabil ψ este construit astfel încât să grupeze descriptori printr-o funcție cost specifică. Aceasta este inspirată din munca lui Wen *et al.* [5] și reduce variațiile intraclasă, încurajând descriptorii să migreze către centrele de clasă corespunzătoare. Folosim o versiune normalizată [7] care limitează zona de migrare. Funcția cost, $\mathcal{L}_{\mathcal{E}}$, poate fi scrisă pe baza descriptorilor normalizați ca $\hat{\mathbf{e}} = \frac{\mathbf{e}}{\|\mathbf{e}\|_2}$:

$$\mathcal{L}_{\mathcal{E}} = \sum_{i=1}^N \left(\mathcal{D}(\hat{\mathbf{e}}_i, \hat{\mathbf{c}}^c) - \frac{1}{C-1} \sum_{j=1, j \neq c}^C \mathcal{D}(\hat{\mathbf{e}}_i, \hat{\mathbf{c}}^j) \right) \quad (1)$$

unde \mathbf{c}^c este centroidul clasei c din totalul claselor C . $\mathcal{D}(\hat{\mathbf{e}}_i, \hat{\mathbf{c}}^j) = \mu_i^c$ arată probabilitatea ca datele \mathbf{x}_i să fie în clasa C . Normalizarea din ecuația (1) limitează spațiul pentru migrația centroidului, în timp ce scăderea impune ca descriptorii să se grupeze către centroizii claselor și să se depărteze de alți centroizi. Centroizii \mathbf{c}^c sunt actualizați după fiecare set astfel:

$$\mathbf{c}^c = \frac{\sum_{i=1}^N \mu_i^c \mathbf{e}_i}{\sum_{i=1}^N \mu_i^c}; \quad \mu_i^c = \begin{cases} 1 & , y_i = c \\ 0 & , y_i \neq c \end{cases} \quad (2)$$

Data: : Subset de b instanțe etichetate în forma one-hot labels cu descriptori
 $\mathcal{X} = \{\dots, (\mathbf{x}_i, \mathbf{e}_i, \mathbf{y}_i), \dots\}, i = 1 \dots b$, parametrul α pentru distribuția β în Eq. (3).

```

for  $b = 1 : N_{batch}$  do
    Calculează descriptorii pentru datele de intrare  $\mathbf{e}_i = \psi(\mathbf{x}_i)$  ;
    Actualizează centrozii folosind ec. (2) ce asigură compactitatea;
    Calculează date sintetice cu ec. (3). ;
    Calculează etichetele pentru date sintetice cu ec. (4). ;

```

end

Calculează costul total cu ec. (5) ;

Actualizează parametrii rețelei;

Algorithm 1: Algoritmul propus SoftClusterMix primește ca intrare un subset de etichete \mathcal{X} și produce exemple sintetice ce populează dens spațiul de intrare \mathcal{X} . Etichetarea noilor date este bazată pe clustering în spațiul descriptorilor. Scopul este să ajustăm ponderile sistemului antrenat ψ .

Date sintetice cu etichete probabilistice. Soluția propusă ține cont de conceptele introduse în MixUp [6], unde, din două instanțe \mathbf{x}_i și \mathbf{x}_j din datele de antrenament, se calculează un nou exemplu, sintetic, utilizând o combinație convexă:

$$\tilde{\mathbf{x}} = \lambda \cdot \mathbf{x}_i + (1 - \lambda)\mathbf{x}_j; \quad \tilde{\mathbf{y}} = \lambda \cdot \mathbf{y}_i + (1 - \lambda)\mathbf{y}_j \quad (3)$$

unde λ este un parametru aleator extras dintr-o distribuție $Beta(\alpha, \alpha)$ de hiperparametru α .

Presupunerea de bază a MixUp este că spațiul de date este structurat astfel încât orice date noi $\tilde{\mathbf{x}}$ să poată fi mai aproape de alte date \mathbf{x}_k și \mathbf{x}_i având eticheta $\tilde{\mathbf{y}} \neq \mathbf{y}_i$ și $\tilde{\mathbf{y}} \neq \mathbf{y}_k$. Considerăm că o astfel de ipoteză poate fi valabilă pentru seturi de date fără exemple etichetate greșit și ușoare, cum ar fi CIFAR-10, dar este invalidă pentru probleme mai dificile.

Pentru a aborda astfel de probleme dificile, noua instanță de date sintetice $\tilde{\mathbf{x}}$ este etichetată în raport cu centrozii claselor. Având în vedere datele sintetice, se calculează descriptorii lui $\tilde{\mathbf{e}}$, cu o metodă inspirată din algoritmul Fuzzy C-means. $\tilde{\mathbf{y}}_c$ este compus din C probabilități de apartenență, iar una dintre ele este:

$$\tilde{\mathbf{y}}_c = 1 / \left(\sum_{j=1}^C \left(\frac{\|\tilde{\mathbf{e}} - \mathbf{c}^c\|_2}{\|\tilde{\mathbf{e}} - \mathbf{c}^j\|_2} \right)^M \right) \quad (4)$$

Parametrul M ajustează dimensiunea vecinătății (cum ar fi gradul de fuzificare a partiției). Acest lucru este echivalent cu luarea în considerare a unei contrastări a probabilităților [1] care compun $\tilde{\mathbf{y}}$.

Costul total. Rețeaua este antrenată folosind costul total calculat ca sumă ponderată de:

$$\mathcal{L} = \lambda_E \cdot \mathcal{L}_E + \lambda_{SM} \cdot \mathcal{L}_{SM} + \lambda_M \cdot \mathcal{L}_M \quad (5)$$

unde λ_E , λ_{SM} și λ_S sunt hiperparametri de ponderare. \mathcal{L}_E este clusterizarea care impune costul pe stratul de descriptori din ecuația (1). \mathcal{L}_{SM} este entropia încrucișată între predicția rețelei și eticheta de referință obținută prin probabilități soft conform ecuației (4). \mathcal{L}_M este entropia încrucișată calculată conform MixUp. Contribuția celor trei pierderi este controlată de ponderile: λ_E și λ_M scad de la 1 la 0 în prima treime a antrenării, în timp ce λ_{SM} crește de la 0 la 1 cu cuante egale pentru fiecare epocă.

Funcția cost este calculată ca medie pe distribuția datelor P

$$R(\psi) = \int \mathcal{L}(\psi(\mathbf{x}), \mathbf{y}) dP(\mathbf{x}; \mathbf{y}) = \frac{1}{N} \sum_{i=1}^N \mathcal{L}_{SM}(\psi(\tilde{\mathbf{x}}), \tilde{\mathbf{y}}) \quad (6)$$

Algoritmul propus: SoftClusterMix Scopul algoritmului este de a antrena o CNN folosind augmentare specifică. Cu toate acestea, înainte de a utiliza structura clusterului pentru a eticheta noile date sintetice, trebuie construită structura în spațiul descriptorilor conform Algoritmului 1.

Tabelul 1: Erori comparative (mai mici este mai bine) pe seturile de date CIFAR-10/100 și SVHN. Valorile sunt obținute cu ResNet-18. Pentru CIFAR, "ERM standard" este preluat din [6], în timp ce pentru SVHN am folosit codul publicat de autor.

Methods	CIFAR-10	CIFAR-100	SVHN
Standard ERM	5.18	25.6	3.81
MixUp [6]	4.24	21.1	2.86
SoftClustMix - proposed	4.85	21.9	3.12

2.1 Experimente

Am evaluat algoritmul propus pe trei baze de date standard: CIFAR-10 și 100 [3], SVHN [4]. Subliniem că aceste baze de date au clase perfect echilibrate. Rezultatele obținute și comparația cu metode relevante sunt în tabelul 1. Se poate observa că rezultatele sunt foarte apropiate de performanța de vârf, cu diferențe neglijabile.

Bibliografie

1. D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, and C. A. Raffel. Mixmatch: A holistic approach to semi-supervised learning. In *NIPS*, pages 5050–5060, 2019.
2. K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
3. A. Krizhevsky. Learning multiple layers of features from tiny images. Technical report, MIT, 2009.
4. Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, and A. Y. Ng. Learning multiple layers of features from tiny images. Technical report, Stanford, 2009.
5. Y. Wen, K. Zhang, Z. Li, and Y. Qiao. A discriminative feature learning approach for deep face recognition. In *ECCV*, pages 499–515, 2016.
6. H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz. mixup: Beyond empirical risk minimization. In *ICLR*, 2018.
7. X. Zhang, Z. Fang, Y. Wen, Z. Li, and Y. Qiao. Range loss for deep face recognition with long-tailed training data. In *CVPR*, pages 5409–5418, 2017.

3 Diseminare

În această etapă am publicat două articole în cadrul conferinței EHB <http://www.ehbconference.ro/Program.aspx>:

1. G.I. Nitroi, C. Florea, V. Pupezescu "Assisted Environment Comprehension using a Mobile Terminal Application"
2. M.C. Dragomir, C. Florea, V. Pupezescu "Automatic Subject Independent Pain Intensity Estimation using a Deep Learning Approach"

Algoritmul prezentat aici este o variantă dezvoltată din soluția semisupervizată publicată în Florea, C., Badea, M., Florea, L., Racoviteanu, A., Vertan, C. "Margin-Mix: Semi-Supervised Learning for Face Expression Recognition" ECCV 2020.

Florea Corneliu