Adversarial reciprocal Point learning

Problem :

n labeled examples $\mathcal{D}_L = \{(x_1, y_1), \ldots (x_n, y_n)\}$

N known classes : $y_i \in \{1, \ldots N\}$

test data : $\mathcal{D}_T = \{t_1, \ldots t_u\}$

t_i label $\in \{1, 2, \ldots N\} \cup \{N+1, \ldots N+U\}$

unknown classes

Deep embedding space for category $k \rightarrow S_k$

& open space $\mathcal{O}_k$ [not k]

$$\mathcal{O}_k = O_k^{pos} \cup O_k^{neg}$$

O_k^{pos} : positive open space of other known classes

O_k^{neg} : infinite unknown space

Here, $\mathcal{D}_L^k \in S^k$

& $\nabla_L^{x_k} \in O_k^{pos}$

$O_u \in O_k^{neg} \Rightarrow \mathbb{R}^d$

Binary prediction function ψ_k (we have one of this for each class N)

ψ_k for k class : 1 ; $x_i \in y_k$

0 ; $x_i \notin y_k$

objective : Learn ψ_k for all known N classes.

$$\underset{\psi_k}{\operatorname{argmin}} \left\{ R^k \;\middle|\; \underset{\in}{R}\left(\psi_k, S_k \cup O_k^{pos}\right) + \alpha \underset{\Downarrow}{R_o}\left(\psi_k, O_k^{neg}\right) \right\}$$

R^k: Expected error

R_o: regularizer

O_k^{neg}: Open risk space

R: Empirical classification risk for known

Further,

$$R_o\left(\psi_k, O_k^{neg}\right) = \frac{\int_{O_k^{neg}} \psi_k(x)\, dx}{\quad}$$

(numerator: should be 0)

$$\int_{S_k \cup \theta_k} \underbrace{\psi_k(x)\, dx}_{\psi \text{ only for } S_k \text{ region}}$$

S_0 for multiclass prediction

$$f = \odot(\psi_1, \psi_2, \cdots \psi_N)$$

↳ integration

final objective:

for all ψ_s

$$\arg\min_{f \in \mathcal{H}} \left\{ \mathcal{R}_\epsilon(f, \mathcal{D}_L) + \alpha \cdot \sum_{k=1}^{N} \mathcal{R}_0(f, \mathcal{D}_U) \right\}$$

$f : \mathbb{R}^d \mapsto N$: measurable function.

Reciprocal point for classification

P^k → latent representation of $\mathcal{D}_L^{\mathcal{F}^k} \cup \mathcal{D}_U$

↳ easier to sample from θ_k than to S_k

$$\max\left(\delta\left(\mathcal{D}_L^{\mathcal{F}^k} \cup \mathcal{D}_U, P^k\right)\right) \leq d ; \forall d \in \delta\left(\mathcal{D}_L^k, P^k\right)$$

↓

set of distance of all sample between the reciprocal point & corresponding known class

Distance between x & the reciprocal point P^k

$$d\left(C(x), P^k\right) = d_e\left(C(x), P^k\right) - d_d\left(C(x), P^k\right)$$

d_e: Euclidean in $\mathbb{R}^m$ (maximize)

d_d: Dot product (- minimize)

[opposite of their RP]

Softmax function for class assignment:

$$P(y=k \mid x, C, P) = \frac{e^{\gamma \overbrace{d\left(C(x), P^k\right)}^{\text{max distance}}}}{\sum_{i=1}^{N} e^{\gamma d\left(C(x), P^i\right)}}$$

Learn θ to minimize:

$$\mathcal{L}_c(x; \theta, \mathcal{P}) = -\log p(y = k \mid \underline{x}, \underline{C}, \underline{\mathcal{P}})$$

↳ corresponding to $\mathcal{R}_\epsilon(f, \mathcal{D}_L)$

We also acheive: $\underset{f \in \mathcal{H}}{\arg\max} \{ \mathcal{G}(\mathcal{D}_L^k, \mathcal{P}^k) \}$

Adversarial margin Constraint:

AMC → minimize $\mathcal{R}_0(f, \mathcal{D}_u)$

Classwise openspace → Global open space

$$\mathcal{O}_G = \bigcap_{k=1}^{N} (\mathcal{O}_k^{pos} \cup \mathcal{O}_k^{neg}) = \bigcap_{k=1}^{N} \mathcal{O}^k$$

further,

$$\max\left(\mathcal{G}(\mathcal{D}_L^{\neq k} \cup \mathcal{D}_u, \mathcal{P}^k)\right) \leq R$$

↓

limiting the open space by R

↙ by

constraint on distance.

$$\mathcal{L}_0(x; \theta, \mathcal{P}^k, R^k) = \max\left(d_E(C(x), \mathcal{P}^k) - R, 0\right)$$

↓

$x \in S^k$

∴ IDEA: explained in th 1

Adverse to each other.

$\mathcal{L}_D$, $\mathcal{L}_E$ are minimize simultaneously if

their exists $\max\left(S\left(D_L^{f^k}, P^k\right)\right) \leq R$

Bound for class k

$$\arg\min_{f \in \mathcal{H}} \left\{ \max\left(\left\{ S\left(D_L^k, P^{f^k}\right) - R \right\} \cup \{0\} \right) \right\}$$

Learning Open set Network:

$$\mathcal{L}(x, y; \Theta, P, R) = \mathcal{L}_c(x; \Theta, P) + \lambda \mathcal{L}_o(x; \Theta, P, R)$$

Adversarial Goal:

$$\max_G \frac{1}{n} \sum_{i=1}^{n} \left[-\frac{1}{N} \sum_{k=1}^{N} \underbrace{S(z_i, P^k)}_{\text{Softmax}\left(d_E\left(C(G(z_i)), P^k\right)\right.} \cdot \log\left(S(z_i, P^k)\right) \right]$$

Empirical ~~Entropy~~ !!

Combined with adversarial setting

GAN Adversse

$$\max_D \quad \log D(x_i) + \log D(1 - G(z_i))$$

$$\&\& \quad \max_G \log D(G(z_i))$$

final Goal:

$$\max_G \left[\log D(G(z_i)) + \beta \, t(z_i, p) \right]$$

GAN — Extra parts.

classifier C is optimized

$$\min_C \frac{1}{n} \sum_{i=1}^{n} \left[L(x_i, y_i) - \beta H(z_i; \Phi) \right]$$

ARPL loss

Reducing open space risk

discriminative.