

Vzdálenostní metriky používané pro měření vzdáleností mezi obrázky

Pavel Jakš

27. července 2023

Úvod

Pod pojmem metrika na prostoru X si každý matematik představí zobrazení $\rho : X \times X \rightarrow [0, +\infty)$ splňující

1. $\rho(x, y) = 0 \iff x = y \quad \forall x, y \in X$,
2. $\rho(x, y) = \rho(y, x) \quad \forall x, y \in X$,
3. $\rho(x, z) \leq \rho(x, y) + \rho(y, z) \quad \forall x, y, z \in X$.

Taková metrika může být na lineárním prostoru V nad číselným tělesem (pro naše účely zůstaňme nad \mathbb{R}) snadno zadána pomocí normy, která je buď indukována skalárním součinem v případě pre-Hilbertových prostorů, nebo dána vlastnostmi, že se jedná o zobrazení $\|\cdot\| : V \rightarrow [0, +\infty)$ a splňuje:

1. $\|x\| = 0 \iff x = 0 \quad \forall x \in V$,
2. $\|\alpha x\| = |\alpha| \cdot \|x\| \quad \forall \alpha \in \mathbb{R}, \forall x \in V$,
3. $\|x + y\| \leq \|x\| + \|y\| \quad \forall x, y \in V$.

Metriku potom získáme z normy následující konstrukcí:

$$\rho(x, y) = \|x - y\|,$$

tedy vzdálenost dvou vektorů je dána normou rozdílu vektorů. Snadno lze nahlédnout, že takto zadané zobrazení je metrika. S metrikami, které jsou tzv. indukované normami dle předchozího se setkáme.

1 Metriky indukované l_p normami

Vzhledem k tomu, že obrázky, které jsou středem naší pozornosti, lze reprezentovat jako tenzory standardně o rozměrech $C \times W \times H$, kde C značí počet kanálů (nejčastěji kanály po řadě pro červenou, zelenou a modrou barvu), W označuje

šířku a H výšku, tak lze na tyto tenzory vpustit L^p normy. Pro $p \in [1, +\infty)$ je L^p norma z $f \in L_p(X, \mu)$ definována vztahem:

$$\|f\|_p = \left(\int_X |f|^p d\mu \right)^{\frac{1}{p}}.$$

Pro naše obrázky lze za X vzít $\{1, \dots, C\} \times \{1, \dots, W\} \times \{1, \dots, H\}$ a za μ *počítací míru*. Potom naše L^p norma přejde v l_p normu, která má pro naše obrázky, tedy tenzory $x \in \mathbb{R}^{C \times W \times H}$, tvar:

$$\|x\|_p = \left(\sum_{i=1}^C \sum_{j=1}^W \sum_{k=1}^H |x_{i,j,k}|^p \right)^{\frac{1}{p}}. \quad (1)$$

Trochu mimo stojí l_∞ norma, která má tvar pro tenzor $x \in \mathbb{R}^{C \times W \times H}$:

$$\|x\|_\infty = \max_{i \in \{1, \dots, C\}} \max_{j \in \{1, \dots, W\}} \max_{k \in \{1, \dots, H\}} |x_{i,j,k}|. \quad (2)$$

A úplně mimo stojí L_0 norma, která svou povahou *není* norma ve smyslu výše uvedené definice, ale pro účely porovnávání obrázků se používá rozdíl obrázků v této pseudo-normě, proto ji zde zmiňuji:

$$\|x\|_0 = |\{x_{i,j,k} \neq 0\}|. \quad (3)$$

2 MSE a RMSE

Vzdálenosti, které mají blízko k metrikám indukovaným l_2 normou, jsou *MSE* (z anglického *Mean Squared Error*) a *RMSE* (z anglického *Root Mean Squared Error*). Pro tenzory $x, \tilde{x} \in \mathbb{R}^{C \times W \times H}$ mají definici:

$$\text{MSE}(x, \tilde{x}) = \frac{1}{CWH} \sum_{i=1}^C \sum_{j=1}^W \sum_{k=1}^H |x_{i,j,k} - \tilde{x}_{i,j,k}|^2 \quad (4)$$

$$\text{RMSE}(x, \tilde{x}) = \left(\frac{1}{CWH} \sum_{i=1}^C \sum_{j=1}^W \sum_{k=1}^H |x_{i,j,k} - \tilde{x}_{i,j,k}|^2 \right)^{\frac{1}{2}} \quad (5)$$

3 Wassersteinova vzdálenost

3.1 Definice

Bud' (M, d) metrický prostor, který je zároveň *Radonův*. Zvolme $p \in [1, +\infty)$. Potom máme *Wassersteinovu p -vzdálenost* mezi dvěma pravděpodobnostními mírami μ a ν na M , které mají konečné p -té momenty, jako:

$$W_p(\mu, \nu) = \left(\inf_{\gamma \in \Gamma(\mu, \nu)} \mathbb{E}_{(x,y) \sim \gamma} d(x,y)^p \right)^{\frac{1}{p}}, \quad (6)$$

kde $\Gamma(\mu, \nu)$ je množina všech sdružených pravděpodobnostních měr na $M \times M$, které mají po řadě μ a ν za marginální pravděpodobnostní míry [1].

Jak to souvisí s obrázky? Přes dopravní problém. Pod pravděpodobnostní distribucí μ či ν na X si lze představit rozložení jakési hmoty o celkové hmotnosti 1. Sdružená rozdělení $\gamma \in \Gamma(\mu, \nu)$ potom odpovídají transportnímu plánu, kde $\gamma(x, y) dx dy$ vyjadřuje, kolik hmoty se přesune z x do y . Tomu lze přiřadit nějakou cenu c , totiž kolik stojí přesun jednotkové hmoty z x do y : $c(x, y)$. V případě *Wassersteinovy vzdálenosti* za cenu dosadíme $c(x, y) = d(x, y)^p$, tedy p -tou mocninu vzdálenosti mezi x a y . Potom cena celkového dopravního problému s transportním plánem γ bude:

$$c_\gamma = \int c(x, y) \gamma(x, y) dx dy \quad (7)$$

$$= \int c(x, y) d\gamma(x, y) \quad (8)$$

a optimální cena bude:

$$c = \inf_{\gamma \in \Gamma(\mu, \nu)} c_\gamma. \quad (9)$$

Po dosazení:

$$c = \inf_{\gamma \in \Gamma(\mu, \nu)} \int c(x, y) d\gamma(x, y) \quad (10)$$

$$= \inf_{\gamma \in \Gamma(\mu, \nu)} \int c(x, y) \gamma(x, y) dx dy \quad (11)$$

$$= \inf_{\gamma \in \Gamma(\mu, \nu)} \mathbb{E}_{(x, y) \sim \gamma} c(x, y) \quad (12)$$

$$= \inf_{\gamma \in \Gamma(\mu, \nu)} \mathbb{E}_{(x, y) \sim \gamma} d(x, y)^p \quad (13)$$

$$= W_p(\mu, \nu)^p \quad (14)$$

Dostáváme tedy interpretaci, že p -tá mocnina *Wassersteinovy vzdálenosti* odpovídá ceně dopravního problému.

Pro obrázky má tato konstrukce následující uplatnění: Obrázky je třeba chápat jako diskrétní pravděpodobnostní rozdělení, proto je třeba je normalizovat, aby součet prvků tenzoru obrázku byl roven 1. Pak střední hodnota v definici *Wassersteinovy vzdálenosti* přejde ve váženou sumu cen, tedy p -tých mocnin vzdáleností mezi jednotlivými pixely.

Jak je to barevnými obrázky, tedy s obrázkem, které mají více než jeden kanál? Zde lze uplatnit následující dva přístupy:

1. Normovat celý obrázek na jedničku, tedy všechny kanály dohromady, a tím pádem i definovat vzdálenost mezi jednotlivými kanály,
2. Normovat každý kanál zvlášť na jedničku, počítat *Wassersteinovu metriku* pro každý kanál zvlášť a následně vybrat nějakou statistiku výsledných vzdáleností, např. průměr.

3.2 Výpočet

Abychom mohli s Wassersteinovou metrikou nakládat například v počítači, je nutné tuto metriku spočítat. Podíváme-li se do definice, znamená to vyřešit optimalizační problém. Byť bychom se omezili hledání vzdáleností dvou vektorů o rozměru q , měli bychom problém s časovou složitostí nejlépe $\mathcal{O}(q^3 \log q)$ [2]. A to je hodně. Proto se podívejme, jak Wassersteinovu vzdálenost spočítat rychleji, byť za ztráty přesnosti.

Omezme se na prostory konečné dimenze. Potom mějme za úkol spočítat Wassersteinovu (zvolme $p = 1$) vzdálenost vektorů $\mu, \nu \in \mathbb{R}^q, \mu^T 1_q = \nu^T 1_q = 1$, kde 1_q je vektor rozměru q složený pouze z jedniček. Potom μ, ν lze chápat jako diskrétní pravděpodobnostní rozdělení. Označme jako $U(\mu, \nu)$ množinu všech matic $P \in \mathbb{R}^{q \times q}, P_{i,j} \geq 0$ takových, že $P 1_q = \mu$ a $P^T 1_q = \nu$. Jako matici C označme zadanou matici cen, která splňuje, že reprezentuje metriku. To znamená, že $C_{i,j} \geq 0, C_{i,j} = 0 \iff i = j, C_{i,j} = C_{j,i}$ a $C_{i,k} \leq C_{i,j} + C_{j,k}$. Potom lze napsat:

$$W(\mu, \nu) \equiv W_1(\mu, \nu) = \min_{P \in U(\mu, \nu)} \langle P, C \rangle, \quad (15)$$

kde $\langle P, C \rangle = \sum_{i,j=1}^q P_{i,j} C_{i,j}$.

Přejdeme nyní od Wassersteinovy metriky k tzv. duální Sinkhornově metrice. Ta je pro pevně zvolené $\lambda > 0$ definována následovně:

$$W^\lambda(\mu, \nu) = \langle P^\lambda, C \rangle, \quad (16)$$

$$\text{kde } P^\lambda = \underset{P \in U(\mu, \nu)}{\operatorname{argmin}} \langle P, C \rangle - \frac{1}{\lambda} H(P), \quad (17)$$

kde $H(P)$ je entropie pravděpodobnostního rozdělení P , tedy

$$H(P) = - \sum_{i,j=1}^q P_{i,j} \log(P_{i,j}).$$

Jedná se tedy o regularizovaný dopravní problém. Tato úprava Wassersteinovy metriky je, jak se přesvědčíme, mnohem lépe vyčíslitelná. Nejdříve se ovšem podívejme na intuici za touto úpravou.

Začneme s mírnou úpravou původního optimalizačního problému definujícího Wassersteinovu vzdálenost: Pro $\alpha > 0$ definujme jakési α okolí rozdělení $\mu\nu^T$ (sdružené pravděpodobnostní rozdělení s marginálními μ a ν , kde μ a ν jsou nezávislá rozdělení) ve smyslu *Kullback-Leiblerovy divergence*

$$U_\alpha(\mu, \nu) = \{P \in U(\mu, \nu) | KL(P || \mu\nu^T) \leq \alpha\}. \quad (18)$$

Připomeňme definici Kullback-Leiblerovy divergence:

$$KL(\tilde{P} || \hat{P}) = \sum_{i,j=1}^q P_{i,j} \log \frac{P_{i,j}}{Q_{i,j}}.$$

Pro dané $P \in U(\mu, \nu)$ lze na kvantitu $KL(P || \mu\nu^T)$ nahlédnout jako na informaci mezi veličinami s rozděleními μ a ν . Tedy $U_\alpha(\mu, \nu)$ vybírá ta rozdělení, která

nesou malou vzájemnou informaci mezi μ a ν (ve smyslu menší než α). Dle [2] lze tuto úpravu ospravedlnit pomocí *principu maximální entropie*.

Potom lze definovat následující Sinkhornovu metriku:

$$W^\alpha(\mu, \nu) = \min_{P \in U_\alpha(\mu, \nu)} \langle P, M \rangle. \quad (19)$$

Jaký je vztah mezi Sinkhornovou metrikou W^α a duální Sinkhornovou metrikou W^λ ? Přes téma duality matematického programování. Zatímco ve W^α figuruje parametr α v omezení definičního oboru, kde optimalizujeme, tak ve W^λ figuruje parametr λ jako Lagrangeův multiplikátor příslušné vazby.

Článek [2] poskytuje též nahlédnutí na fakt, že W^λ a W^α jsou skutečně metriky.

Tento úrok stranou pomocí entropické regularizace původního problému lineárního programování, jehož vyřešení je nutné pro výpočet Wassersteinovy vzdálenosti, poskytuje úlevu v oblasti časové složitosti pro výpočet.

4 PSNR

Vzdálenost označená zkratkou *PSNR* z anglického *Peak Signal-to-Noise Ratio* vyjadřuje vztah mezi obrázkem $x \in \mathbb{R}^{C \times W \times H}$ a jeho pokažením $\tilde{x} \in \mathbb{R}^{C \times W \times H}$ za přidání šumu. Definice je následující:

$$\text{PSNR}(x, \tilde{x}) = 10 \cdot \log_{10} \left(\frac{l^2}{\text{MSE}(x, \tilde{x})} \right), \quad (20)$$

$$= 20 \cdot \log_{10} \left(\frac{l}{\text{RMSE}(x, \tilde{x})} \right), \quad (21)$$

kde l je dynamický rozsah obrázků, tedy rozdíl mezi maximální možnou hodnotou pixelů a minimální možnou hodnotou pixelů. Jedná se tedy o transformaci metriky *MSE*.

5 SSIM

Zkratka *SSIM* pochází z anglického *structural similarity index measure*. Tato metrika se při výpočtu indexu dvou obrázků x a \tilde{x} dívá na podokna, ze kterých vybere jisté statistiky a z nich vytvoří index pro daná podokna obrázků. Potom se jako celkový index bere průměr přes tato okna. Uveďme vzorce pro výpočet indexu SSIM pro případ, že máme jediné okno, které splývá s obrázkem, které pro jednoduchost zvolme jednokanálové, tedy černobílé. Označme $N = W \times H$ počet pixelů v obrázku a indexujme prvky matice obrázku jediným číslem.

Potom definujeme pro obrázky x a \tilde{x} následující:

$$\begin{aligned}\mu_x &= \frac{1}{N} \sum_{i=1}^N x_i, \\ \mu_{\tilde{x}} &= \frac{1}{N} \sum_{i=1}^N \tilde{x}_i, \\ \sigma_x^2 &= \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)^2, \\ \sigma_{\tilde{x}}^2 &= \frac{1}{N-1} \sum_{i=1}^N (\tilde{x}_i - \mu_{\tilde{x}})^2, \\ \sigma_{x\tilde{x}} &= \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)(\tilde{x}_i - \mu_{\tilde{x}}).\end{aligned}$$

Potom:

$$\text{SSIM}(x, \tilde{x}) = \frac{(2\mu_x\mu_{\tilde{x}} + C_1)(2\sigma_{x\tilde{x}} + C_2)}{(\mu_x^2 + \mu_{\tilde{x}}^2 + C_1)(\sigma_x^2 + \sigma_{\tilde{x}}^2 + C_2)}, \quad (22)$$

kde C_1, C_2 jsou konstanty pro stabilitu dělení volené kvadraticky úměrně dynamickému rozsahu.

Jak volíme celkový SSIM pro barevné obrázky? Jako průměr přes kanály.

Reference

- [1] L. Vaserstein, *Markov processes over denumerable products of spaces, describing large systems of automata*. Problemy Peredači Informacii 5, 1969.
- [2] M. Cuturi, *Sinkhorn Distances: Lightspeed Computation of Optimal Transport*. Advances in Neural Information Processing Systems 26, 2013.