

Métricas de Evaluación de IA Generativa

Lenin G. Falconí

2024-10-07

Índice

| | |
|---|----------|
| 1. TODO Instrucciones [100 %] | 1 |
| 2. Frechet Inception Distance (FID) | 1 |
| 2.1. Concepto | 1 |
| 2.2. Ecuación: | 2 |
| 2.3. Interpretación | 2 |
| 2.4. Limitaciones de la métrica | 2 |
| 2.5. Ejemplo de cálculo | 2 |
| 3. Learned Perceptual Image Patch Similarity (LPIPS) | 2 |
| 3.1. Concepto: | 2 |
| 3.2. Ecuación | 3 |
| 3.3. Interpretación | 3 |
| 3.4. Limitaciones de la métrica | 3 |
| 3.5. Ejemplo de cálculo | 3 |
| 4. Aplicaciones Prácticas | 3 |

1. **TODO Instrucciones [100 %]**

- ☒ Investigar las métricas asignadas FID: Frechet Inception Distance
- ☒ Investigar las métricas asignadas LPIPS: Learned Perceptual Image Patch Similarity
- ☒ Definir la métrica y su fórmula básica
- ☒ ¿Cómo se interpreta la métrica y qué representa en el contexto específico?
- ☒ Ventajas y Limitaciones de la métrica
- ☒ Indicar cómo se puede aplicar la métrica en un caso específico
- ☒ Describir el tipo de dato a utilizar
- ☒ Detallar los pasos para calcular la métrica

2. **Frechet Inception Distance (FID)**

2.1. **Concepto**

- Calcula la distancia entre los feature vectors calculados de imágenes reales y generadas [1].
- Indica que tan similares son dos grupos en términos estadísticos de la media μ y la covarianza Σ de componentes de *computer vision*
- Scores más bajos indican que los grupos son más similares.
- Evalúa la calidad de las imágenes generadas [3].

- Propuesto en [4] en el año 2017.
- Técnicamente utiliza la última capa de pooling de la red Inception V3 de 2048 componentes [3].

2.2. Ecuación:

La Ecuación 1 define el cálculo de FID como:

$$FID = \|\mu - \mu_w\|^2 + \text{tr}(\Sigma + \Sigma_w - 2(\Sigma\Sigma_w)^{\frac{1}{2}}) \quad (1)$$

Donde:

$\mathcal{N}(\mu, \Sigma)$ distribución multivariable normalizada de un inception v3 sobre imágenes reales

$\mathcal{N}(\mu_w, \Sigma_w)$ distribución multivariable normalizada de un inception v3 sobre imágenes generadas (falsas)

2.3. Interpretación

La fórmula del FID combina la diferencia en las medias ($(\|\mu - \mu_w\|^2)$) y la diferencia en las covarianzas ($(\text{tr}(\Sigma + \Sigma_w - 2(\Sigma\Sigma_w)^{\frac{1}{2}}))$) de las características extraídas de las imágenes reales y generadas. Un **FID más bajo** indica que las imágenes generadas **son más similares** a las imágenes reales en términos de sus características estadísticas.

2.4. Limitaciones de la métrica

En [2] se demuestra que tanto el *Inception Score* como la Frechet Inception Distance son métricas sesgadas. El valor obtenido por ambas métricas se demuestra no confiable. Pues, su valor depende del sesgo del Modelo y no de la métrica en sí. A fin de corregir este problema, se sugiere la extrapolación de las métricas usando métodos de integración de Quasi-Monte Carlo.

2.5. Ejemplo de cálculo

La métrica está contenida dentro de la librería de Pytorch. Se utiliza el extractor de características de Inception v3. Los datos de entrada son mini batches de imágenes con 3 canales RGB de forma $(3 \times H \times W)$. Las imágenes se redimensionan a 299×299 .

El siguiente código permite calcular Frechet Inception Distance utilizando la capa 64 del modelo de Inception. Para esto se genera dos muestras de 100 imágenes. Las imágenes reales varían aleatoriamente sus valores entre 0 y 199 y las falsas entre 100 y 255. En ambos casos las imágenes son de 3 canales (RGB) con un tamaño de 299×299

```
import torch
from torchmetrics.image.fid import FrechetInceptionDistance
fid = FrechetInceptionDistance(feature=64)
imgs_dist1 = torch.randint(0, 200, (100, 3, 299, 299), dtype=torch.uint8)
imgs_dist2 = torch.randint(100, 255, (100, 3, 299, 299), dtype=torch.uint8)
fid.update(imgs_dist1, real=True)
fid.update(imgs_dist2, real=False)
print(fid.compute())
```

tensor(12.6172)

3. Learned Perceptual Image Patch Similarity (LPIPS)

3.1. Concepto:

- Calcula la similitud perceptiva de dos imágenes [5].
- Computa la similitud entre las activaciones de dos patches imágenes para una red convolucional pre definida [5].
- Iguala la percepción humana.

- Un score más bajo de LPIPS implica que las imágenes son más semejantes.
- Propuesto en el paper [6] de 2018.
- De acuerdo a los autores en [5], las activaciones internas de redes entrenadas en tareas de clasificación de alto nivel (e.g. ImageNet, VGG) se corresponden con el juicio de percepción humana.

3.2. Ecuación

Dadas una imagen real x y una imagen generada x_0 con una red neuronal convolucional \mathcal{F} , se obtienen la pila de vectores componentes (*feature stack*) de \mathcal{L} capas, y se normalizan con respecto a la dimensión del canal \hat{y}^l , \hat{y}_0^l para una capa l . Las activaciones se escalan con respecto del vector w^l y se calcula la distancia ℓ_2 , como indica la ecuación 2

$$d(x, x_0) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} \|w_l \odot (\hat{y}_{hw}^l - \hat{y}_{0hw}^l)\|_2^2 \quad (2)$$

3.3. Interpretación

La métrica se basa en la hipótesis de que una red pre-entrenada para resolver una tarea de clasificación visual contiene *deep embeddings* que son de interés y emulan la percepción visual humana. Para esto, se obtiene las componentes normalizadas de las imágenes x y x_0 en una capa l (el paper considera las capas 1 a 5). Luego se obtiene la normalización euclidia ℓ_2 normalizada.

3.4. Limitaciones de la métrica

3.5. Ejemplo de cálculo

Los image patches a ser utilizados han de tener la forma (N,3,H,W). El siguiente código presenta cómo se puede utilizar la librería de `torchmetrics` para obtener LPIPS en una muestra de 10 imágenes RGB de 100×100

```
from torch import rand
from torchmetrics.image.lpip import LearnedPerceptualImagePatchSimilarity
lpips = LearnedPerceptualImagePatchSimilarity(net_type='squeeze')
img1 = (rand(10, 3, 100, 100) * 2) - 1
img2 = (rand(10, 3, 100, 100) * 2) - 1
print(lpips(img1, img2))

tensor(0.1017)
```

4. Aplicaciones Prácticas

En ambos casos, tanto FID como LPIPS se pueden utilizar para evaluar la calidad de imágenes generadas con un modelo generativo. Una aplicación, sería la generación de imágenes médicas sintéticas. Por ejemplo, las imágenes de termografía de cáncer de mama suelen no ser muy frecuentes dado que la mamografía se considera médicamente como el gold standard. La generación de imágenes de buena calidad puede ayudar a explorar modelos de deep learning aplicados a diagnóstico asistido por computador.

Referencias

- [1] Jason Brownlee. *How to Implement the Frechet Inception Distance (FID) for Evaluating GANs - MachineLearningMastery.com* — *machinelearningmastery.com*. [Accessed 09-10-2024]. 2019. URL: <https://machinelearningmastery.com/how-to-implement-the-frechet-inception-distance-fid-from-scratch/>.
- [2] Min Jin Chong y David Forsyth. “Effectively Unbiased FID and Inception Score and where to find them”. En: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (nov. de 2019), págs. 6069-6078. ISSN: 10636919. DOI: 10.1109/CVPR42600.2020.00611. URL: <https://arxiv.org/abs/1911.07023v3>.

- [3] Wikipedia contributors. *Fréchet inception distance*. [Online; accessed 9-October-2024]. 2024. URL: https://en.wikipedia.org/w/index.php?title=Fr%C3%A9chet_inception_distance&oldid=1250059515.
- [4] Martin Heusel et al. “GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium”. En: *Advances in Neural Information Processing Systems* 2017-December (jun. de 2017), págs. 6627-6638. ISSN: 10495258. DOI: 10.18034/ajase.v8i1.9. URL: <https://arxiv.org/abs/1706.08500v6>.
- [5] lightning.ai. *Learned Perceptual Image Patch Similarity (LPIPS)*. https://lightning.ai/docs/torchmetrics/stable/image/learned_perceptual_image_patch_similarity.html. [Accessed 09-10-2024]. 2024.
- [6] Richard Zhang et al. “The Unreasonable Effectiveness of Deep Features as a Perceptual Metric”. En: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (ene. de 2018), págs. 586-595. ISSN: 10636919. DOI: 10.1109/CVPR.2018.00068. URL: <https://arxiv.org/abs/1801.03924v2>.