УДК 004.7 **doi: 10.26906/SUNZ.2025.1.134-137**

А. В. Шостак

Національний аерокосмічний університет «ХАІ», Харків, Україна

ПРО ОСОБЛИВОСТІ ФОРМУВАННЯ ТРИПЛЕТІВ ДЛЯ НАВЧАННЯ СІАМСЬКОЇ НЕЙРОННОЇ МЕРЕЖІ

Анотація. **Предмет дослідження** – процеси розпізнавання зображень із застосуванням нейронних мереж. Додаток для розпізнавання ґрунтується на архітектурі сіамської мережі з триплетною функцією втрат і зі згортковою нейронною підмережею. **Мета статті** – обґрунтування вибору квазівипадкової N-вимірної послідовності векторів як вкладень для навчання сіамської нейронної мережі з триплетною функцією втрат та оцінка отриманих після навчання характеристик кластерів вкладень зображень цифр. **Завдання**: експериментальна оцінка отриманих після навчання сіамської нейронної мережі з триплетною функцією втрат характеристик кластерів вкладень зображень цифр. **Методи досліджень**: метод прямого пошуку для функцій з кількома змінними для визначення оцінок N-вимірних векторних представлень вхідних зображень. **Результати досліджень**. Проведено дослідження квазівипадкової N-вимірної послідовності векторів як вкладень для навчання сіамської нейронної мережі з триплетною функцією втрат та її тестування. Показано, що запропоновані методи визначення N-вимірних векторних представлень вхідних зображень є робастними та значно зменшують обсяг обчислень під час навчання. Під час тестування використовувалися зображення рукописних цифр із тестового набору MNIST. Визначено, що використання квазівипадкової N-вимірної послідовності векторів як вкладень зображень під час навчання сіамської нейронної мережі з триплетною функцією втрат дає змогу значно поліпшити характеристики отриманих кластерів вкладень зображень. **Висновки**. Результати, отримані в роботі, можуть бути використані при порівняльній оцінці та визначенні N-вимірних векторних представлень різних класів вхідних зображень з метою їх розпізнавання з використанням архітектури сіамської нейронної мережі з триплетною функцією втрат.

Ключові слова: сіамська нейронна мережа, триплет, вкладення, прототип вкладень кластера, тестування нейронної мережі.

Вступ

Сіамська нейронна мережа (СНМ) – це один із видів нейронної мережі, яка широко використовується в системах розпізнавання обличь та інших графічних образів, у системах перевірки підпису, для порівняння текстів тощо [1–3].

Однією з найефективніших під час навчання СНС є триплетна функція втрат такого вигляду [4]:

$$\text{Loss} = \sum_{i=1}^{k} \text{L}((\text{A}, \text{P}, \text{N})_i), \quad (1)$$

причому

$$\text{L}((\text{A}, \text{P}, \text{N})_i) = \max\{d(\text{A}, \text{P})_i - d(\text{A}, \text{N})_i + \text{margin}, 0\} –$$

величина втрат для і-го триплету; k – розмір набору триплетів використовуваних для навчання мережі; $(\text{A}, \text{P}, \text{N})_i$ – і-й триплет зображень, що складається з якірного зображення A, позитивного зображення P (зображення A і P належать до одного класу зображень) і негативного зображення N (зображення A і N належать до різних класів зображень), $d(\text{A}, \text{P})_i$ – евклідова відстань між вкладеннями і-й пари зображень A та P, $d(\text{A}, \text{N})_i$ – евклідова відстань між вкладеннями і-й пари зображень A та N, margin > 0 – параметр.

Якірні зображення A – це дані деякого класу, які визначають, на якому класі триплет буде навчати СНС модель.

Мета використання триплетної функції втрат - мінімізувати відстань між вкладеннями зображень A і P, водночас максимізуючи відстань між вкладеннями A і N, тобто зробити кластери вкладень зображень більш компактними, такими, що не перетинаються, та щоб кластери різних класів були на максимальній відстані один від одного.

Відповідно до виразу для триплетної функції втрат виділяють три види триплетів [4]:

– легкі триплети, які мають втрату, що дорівнює 0, тому що $d(\text{A}, \text{P})_i + \text{margin} < d(\text{A}, \text{N})_i$;

– жорсткі триплети, в яких вкладення негативного зображення N ближче до вкладення якоря A, ніж вкладення позитивного зображення P, тобто $d(\text{A}, \text{N})_i < d(\text{A}, \text{P})_i$;

– напівжорсткі триплети, в яких вкладення негативного зображення N не ближче до вкладення якоря A, ніж вкладення позитивного зображення P, але значення втрати все ще залишається позитивним $d(\text{A}, \text{P})_i < d(\text{A}, \text{N})_i < d(\text{A}, \text{P})_i + \text{margin}$.

Зазвичай, для оптимізації якості та швидкості навчання СНС, набори триплетів із заданими властивостями (y, x, z) (де y - кількість жорстких триплетів у наборі, x - кількість легких триплетів, z - кількість напівжорстких триплетів, k - розмір набору триплетів, причому x + y + z = k) і вкладення для їхніх складових формують безпосередньо під час навчання. Так, наприклад, у [4] k =256, y = k\2, x+z = k\2 (причому співвідношення x і z між собою чітко не задається). У [5] k = 32, y = k\2, x+z = k\2 (співвідношення x і z між собою також чітко не задають), причому набір із k триплетів формують із випадкового набору з 200 триплетів. У [6] k = 256 і гарантується, що щонайменше z = k\2, а решта триплетів можуть бути довільними. Тобто відповідно до [4, 5, 6] жорсткі та напівжорсткі триплети забезпечують більш якісне навчання СНС. Такий підхід до структури і складу триплетів видається не цілком ефективним з погляду обсягу обчислень, якості та швидкості навчання СНС.

Основна частина

Оскільки у функції втрат (1) використовуються евклідові відстані між вкладеннями пари зображень A та P та пари зображень A та N, то під час навчання СНС обчислюватимемо вкладення якірних

зображень A та використовуватимемо оцінки прототипів вкладень класів зображень EP та EN. Тоді вираз для триплетних втрат матиме вигляд

$L((A, EP, EN)_i) = \max\{d(A, EP)_i - d(A, EN)_i + margin, 0\}$.

Тут зображення A і вкладення EP належать до одного класу, зображення A і вкладення EN належать до різних класів, а вкладення EP і EN є оцінками прототипів вкладень різних класів зображень.

Як оцінки прототипів вкладень EP і EN під час навчання СНС було використано квазівипадкову n-вимірну (n=10) послідовність із 10 векторів, генерованих за методом Соболя [3, 7].

Структуру підмережі СНС [5], що формує вкладення для вхідних зображень, наведено на рис. 1. На вхід підмережі СНС подається з набору даних MNIST [8] одноканальне зображення у градаціях сірого кольору з рукописною цифрою розміру 28*28 пікселів - Input(28*28*1). Далі йде згортковий шар Conv2D_1(128, (7, 7), relu)), який формує за допомогою ядер згортки розміру (7, 7) 128 карт ознак і використовує функцію активації relu. Після першого шару згортки слідує шар підвибірки MaxPooling2D, що замінює дані у вікні їхнім максимальним значенням. Далі - другий шар згортки Conv2D_2(128, (3, 3), relu) з 128 картами ознак, шар підвибірки MaxPooling2D і третій шар згортки Conv2D_3(256, (3, 3), relu). Потім шар Flatten, на виході якого з його вхідних даних формується одновимірний вектор. Далі повнозв'язний шар Dense(4096, relu) з 4096 вузлів з використанням функції активації relu і повнозв'язний шар Dense(10) з n=10 вузлів без використання функції активації для забезпечення повного діапазону значень вкладення. Вихід підмережі формується Lambda шаром, який виконує операцію L2-нормалізації значень вкладення з попереднього шару Dense(10).

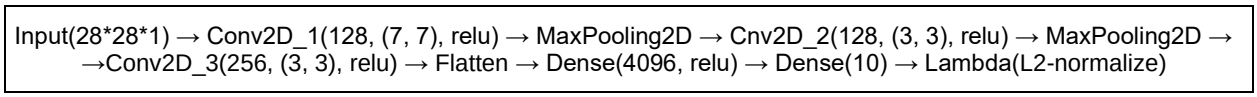
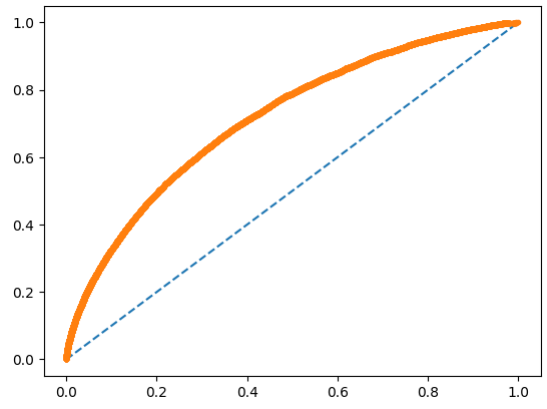


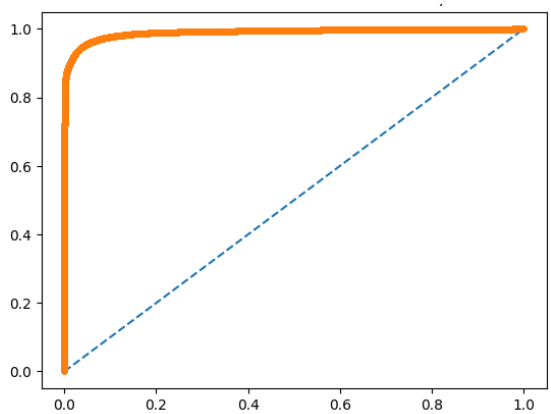
Рис. 1. Структура підмережі СНС

На кожній ітерації навчання формувався набір із 32 триплетів у такий спосіб. Спочатку на підставі тренувального набору і вкладень для десяти цифр генерувався набір із 200 випадкових триплетів виду $(A, EP, EN)_i$. Потім цей набір сортувався за зменшенням різниці між відстанями $d(A, EP)_i$ і $d(A, EN)_i$ і з відсортованого набору вибирали 16 перших триплетів. Далі з решти набору з 200 триплетів вибирали ще 16 довільних триплетів. Модель було навчено на 9000 ітерацій. Для оцінювання якості моделі СНС на рис. 2 представлено розраховану на підставі вкладень тестового набору зображень ROC-криву [9] а) до навчання, б) після навчання, а також значення площі під кривою (AUC), чутливості (Sensitivity) і порогу (Threshold), який забезпечує частоту хибнопозитивних результатів (FPR) не більше ніж 10e-3.



AUC = 0.717, Sensitivity = 1.9%, Threshold = 0.0448

а



AUC = 0.990, Sensitivity = 80.1%, Threshold = 0.2216

б

Рис. 2. ROC-крива: а – до навчання, б – після навчання

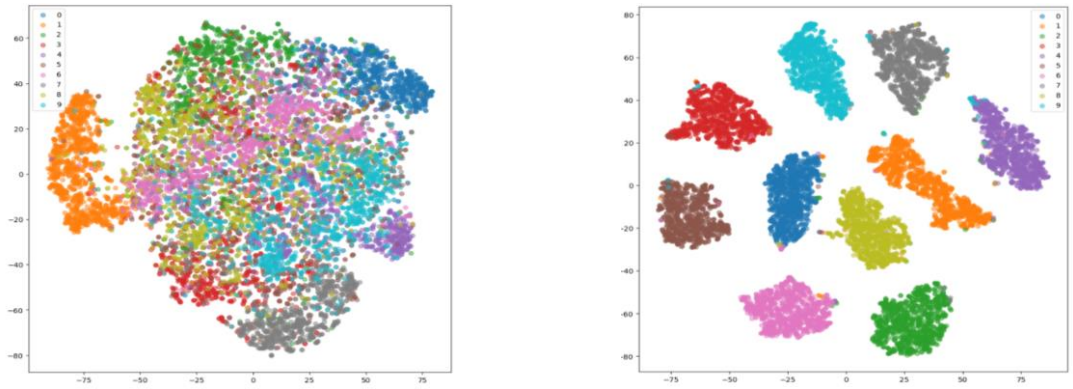
На рис. 3 показано візуалізовані з використанням алгоритму зниження розмірності t-SNE [10] десять кластерів вкладень рукописних цифр тестового набору зображень MNIST а) до навчання та б) після навчання.

Бачимо, що після навчання з використанням триплетної функції втрат, триплетів вигляду $(A, EP, EN)_i$ і підмережі СНС (рис. 1), усі десять кластерів вкладень практично не перетинаються, і відстані між вкладеннями кожного кластера стали істотно меншими.

Прототипи вкладень кластерів вкладень зображень цифр є узагальненою характеристикою кластерів вкладень відповідних цифр і використовуються для розрахунку ступеня схожості вхідного зображення із зображенням відповідної цифри.

Для пошуку прототипів вкладень зображень цифр тестового набору даних було використано метод Пауелла прямого пошуку для функції з кількома змінними з Python-бібліотеки SciPy. Прототипи вкладень зображень цифр для підмережі СНС (рис. 1) після навчання мають вигляд:

0) [-0.2824016 -0.27252399 -0.37384685 -0.33512156 -0.29360956 -0.30970557 -0.27883729 -0.33485432 -0.33369731 -0.3317796];
1) [-0.05223146 -0.39773643 0.1311696 -0.11012863 -0.00831585 -0.24040615 0.46429436 0.28060366 0.45396905 0.13724699];
2) [0.29550691 -0.24569951 -0.32850522 -0.30957372 0.33073034 0.32076124 -0.20231111 0.33361188 0.34647749 0.36148253];
3) [-0.29281085 0.31785711 0.28628017 0.36124768 -0.35524327 -0.27347703 0.31099376 -0.32188501 -0.32414977 -0.25594271];
4) [-0.2123403 -0.13414584 0.13346918 0.29032724 -0.09325196 -0.50832853 -0.15217541 0.50219026 0.4942752 0.1325437];
5) [0.37764855 0.40513021 -0.38194006 -0.13066819 0.39994588 0.11103943 0.40234225 -0.10447713 -0.10416116 -0.40455939];
6) [0.10466758 -0.38080509 0.42403376 0.16230785 0.11301583 0.3711326 -0.39282424 -0.39239505 -0.3939226 -0.12383774];
7) [-0.40467949 0.18668864 -0.1246697 -0.40202019 -0.38693982 -0.13544805 0.19201151 0.25637364 0.26939899 0.48835121];
8) [-0.35404521 -0.19320955 0.4521353 -0.07188427 0.08048192 -0.25373991 -0.09305534 0.48344564 0.49281425 -0.21215699];
9) [0.17273745 0.34933134 -0.0630242 0.45769747 -0.45803124 0.2907892 0.46623946 -0.04591382 -0.03500779 0.32947503].



а – до навчання б – після навчання

Рис. 3. Візуалізація десяти кластерів вкладень рукописних цифр тестового набору зображень MNIST

У табл. 1 наведено розраховані для тестового набору даних розміру 10000 з використанням прототипів вкладень характеристики кластерів вкладень зображень цифр - кількість зображень різних цифр у тестовому наборі (k), мінімальна відстань (r), максимальна відстань (радіус кластера) між вкладеннями кластера і прототипом вкладення (R), середня внутрішньокластерна відстань (SSR) і середнє квадратичне відхилення (sD). У табл. 1 перший рядок для кожної цифри відповідає стану СНС до навчання, другий рядок - після 9000 ітерацій навчання.

Таблиця 1 – **Характеристики кластерів вкладень зображень цифр**

Цифра	k	r	R	SSR	sD
0	980	0.0769	0.7303	0.3171	0.0961
		0.0043	1.1589	0.0287	0.0517
1	1135	0.0889	0.8783	0.2924	0.1082
		0.1453	1.4764	0.5182	0.2003
2	1032	0.1900	0.8640	0.4255	0.1078
		0.0454	1.4558	0.2137	0.1666
3	1010	0.09336	0.7289	0.3342	0.0905
		0.04274	1.8494	0.2048	0.1668
4	982	0.1466	0.8356	0.3705	0.1088
		0.0292	1.3727	0.2093	0.1615
5	892	0.1239	0.6659	0.3461	0.0913
		0.0218	1.5957	0.1306	0.1513
6	958	0.1645	0.7195	0.3818	0.0970
		0.0192	1.7079	0.1220	0.1641
7	1028	0.1058	0.7845	0.3298	0.0991
		0.0439	1.1062	0.2406	0.1536
8	974	0.1344	0.7767	0.3628	0.0930
		0.0344	1.4964	0.1966	0.1710
9	1009	0.0883	0.6607	0.3067	0.0874
		0.0274	1.4920	0.1363	0.1614

Відповідно до даних із табл. 1 до навчання найбільш некомпактний за показником R кластер вкладень зображень цифри 1 і кластер вкладень зображень цифри 2 за показником SSR. Найкомпактніший за показником R кластер вкладень зображень цифри 9 і кластер вкладень зображень цифри 1 за показником SSR.

Після навчання найбільш некомпактний за показником R кластер вкладень зображень цифри 6 і кластер вкладень зображень цифри 1 за показником SSR. Найкомпактніший за показником R кластер вкладень зображень цифри 7 і кластер вкладень зображень цифри 0 за показником SSR.

Після навчання значення мінімальної відстані r для кластерів вкладень зображень усіх цифр, крім цифри 1, зменшилися від 2,2 до 17,9 разів. Для вкладень цифри 1 значення r після навчання збільшилися в 1,6 раза. Значення максимальної відстані R для всіх кластерів вкладень цифр збільшилися від 1,4 до 2,5 разів. Значення середньої внутрішньокластерної відстані SSR для всіх кластерів вкладень цифр, крім цифри 1, зменшилися від 1,4 до 11,0 разів. Для вкладень цифри 1 значення SSR після навчання збільшилися в 1,8 разів. Значення середнього квадратичного відхилення sD для всіх кластерів вкладень цифр, крім вкладень цифри 0, збільшилися від 1,5 до 1,9 раза. Для вкладень цифри 0 значення sD після навчання зменшилися в 1,9 раза.

У табл. 2 для трьох 10-вимірних наборів вкладень розміру n=10 (перший набір - вектори квазівипадкової послідовності Соболя, другий набір - прототипи вкладень кластерів рукописних цифр для моделі СНС (рис. 1) до навчання і третій набір - прототипи вкладень кластерів рукописних цифр для моделі СНС (рис. 1) після навчання) наведені відповідно мінімальна відстань (rE) між вкладеннями в наборі, максимальна відстань (RE) і середня відстань (SSRE).

У результаті навчання на підставі розрахунків на тестовій вибірці (рядки 2 і 3 табл. 2) мінімальна відстань rE між прототипами вкладень кластерів рукописних цифр збільшилася в 3,26 рази, максимальна

відстань RE - у 2,63 раза і середня відстань SSRE - у 4,00 раза.

Таблиця 2 – **Характеристики трьох наборів вкладень**

N	rE	RE	SSRE
1	1.4252	4.3768	2.6435
2	0.2112	0.7458	0.3536
3	0.6885	1.9595	1.4129

Висновки

У роботі досліджено спосіб використання розрахованої за методом Соболя квазівипадкової n-вимірної послідовності векторів для формування триплетів під час навчання СНС.

Використання квазівипадкової n-вимірної послідовності векторів як вкладень EP і EN триплету (A, EP, EN) усуває необхідність застосування позитивного і негативного зображень P і N та обчислення підмережею СНС вкладень для них. Тобто завдяки використанню триплету виду (A, EP, EN) значно скорочується обсяг обчислень під час навчання СНС. При цьому в результаті навчання СНС на триплетах виду (A, EP, EN) за 9000 ітерацій площа під ROC-кривою збільшилася з 0,717 до 0,99, чутливість - з 1,9% до 80,1% і поріг, що забезпечує частоту хибнопозитивних результатів не більше 10e-3, з 0,045 до 0,222.

Ефект від навчання СНС із триплетною функцією втрат і з триплетом виду (A, EP, EN) полягає в значному відокремленні та стисненні кластерів вкладень зображень рукописних цифр. Так після навчання значення мінімальної відстані для кластерів вкладень зображень усіх цифр тестової вибірки, окрім цифри 1, зменшилися від 2,2 до 17,9 разів. Значення середньої внутрішньокластерної відстані для всіх кластерів вкладень цифр, крім цифри 1, зменшилися від 1,4 до 11,0 разів. Також у результаті навчання мінімальна відстань між прототипами вкладень кластерів цифр збільшилася в 3,26 рази, максимальна відстань - у 2,63 рази і середня відстань - у 4,00 рази.

Подальші дослідження слід спрямувати на вибір адаптивних алгоритмів перерахунку векторів вкладень для формування триплетів під час навчання з метою прискорення навчання і збільшення його якості, а також для поліпшення характеристик кластерів вкладень зображень.

Список літератури

1. Chicco D. Siamese Neural Networks: An Overview. Artificial Neural Networks. MIMB, vol. 2190, 2020, pp. 73-94. URL: https://link.springer.com/protocol/10.1007/978-1-0716-0826-5_3
2. Шостак А. В. Про особливості формування дескрипторів у сіамській нейронній мережі. Системи управління, навігації та зв'язку, Полтава: НУ ПП, 2021, вип. 4(66). С. 91-96. DOI: https://doi.org/10.26906/SUNZ.2021.4.079
3. Шостак А. В. Про особливості формування вхідних даних у сіамській нейронній мережі. Системи управління, навігації та зв'язку, Полтава: НУ ПП, 2024, вип. 3(77). С. 193-195. DOI: 10.26906/SUNZ.2024.3.193
4. Schroff F., Kalenichenko D., Philbin J. FaceNet: A unified embedding for face recognition and clustering. Proceedings of the IEEE CSC on CVPR, 2015, pp. 815-823.
5. Craeymeersch E. One Shot learning, Siamese networks and Triplet Loss with Keras. URL: https://medium.com/@crimy/one-shot-learning-siamese-networks-and-triplet-loss-with-keras-2885ed022352
6. Trotter C. How To Train Your Siamese Neural Network. URL: https://towardsdatascience.com/how-to-train-your-siamese-neural-network-4c6da3259463
7. Owen A.B. On Dropping the First Sobol' Point. In: Keller, A. (eds) Monte Carlo and Quasi-Monte Carlo Methods. MCQMC 2020. Springer Proc. in Mathematics & Statistics, vol 387. Springer, Cham. DOI: https://doi.org/10.1007/978-3-030-98319-2_4
8. The Mnist database of handwritten digits. URL: http://yann.lecun.com/exdb/mnist/
9. Hernandez-Orallo J. ROC curves for regression. Pattern Recognition. 2013. vol. 46, no. 12. pp. 3395–3411. doi: 10.1016/j.patcog.2013.06.014
10. Van der Maaten L.J.P. Accelerating t-SNE using Tree-Based Algorithms. Journal of Machine Learning Research 15(Oct). – 2014. - С. 3221-3245.

Received (Надійшла) 15.12.2024
Accepted for publication (Прийнята до друку) 05.03.2025

On the features of triplet's formation for Siamese neural network training

A. Shostak

Анотація. Summary. The subject of research – image recognition processes using neural networks. The recognition application is based on a Siamese network architecture with a triplet loss function and a convolutional neural subnetwork. **The purpose of the article** – to justify the choice of a quasi-random N-dimensional sequence of vectors as embeddings for training a Siamese neural network with a triplet loss function and to evaluate the characteristics of the clusters of digit image embeddings obtained after training. **Objective**: an experimental evaluation of the characteristics of clusters of embeddings of digit images obtained after training a Siamese neural network with a triplet loss function. **Research methods**: a direct search method for functions with several variables to determine estimates of N-dimensional vector representations of input images. **Research results**. A study of a quasi-random N-dimensional sequence of vectors as embeddings for training a Siamese neural network with a triplet loss function and its testing is carried out. It is shown that the proposed methods for determining N-dimensional vector representations of input images are robust and significantly reduce the amount of computation during training. During testing, images of handwritten digits from the MNIST test set were used. It has been determined that the use of a quasi-random N-dimensional sequence of vectors as image embeddings in training a Siamese neural network with a triplet loss function can significantly improve the characteristics of the obtained image embedding clusters. **Conclusions**. The results obtained in this work can be used for comparative evaluation and determination of N-dimensional vector representations of different classes of input images for their recognition using the architecture of a Siamese neural network with a triplet loss function.

Ключові слова: Siamese neural network, triplet, embedding, prototype cluster embedding, neural network testing.