소량의 필적 샘플을 활용한 Vision Transfomer 기반 메타러닝 작성자 식별 기법

김인곤, 신수용*

국립금오공과대학교

20246104@kumoh.ac.kr, *wdragon@kumoh.ac.kr

Meta-Learning-Based Writer Identification Using Vision Transformers with Limited Handwriting Samples

In Gon Kim, Soo Young Shin*

Kumoh National Institute of Technology.

요 약

본 연구에서는 손글씨 샘플을 기반으로 작성자의 글씨체를 식별하는 메타러닝 모델을 제안한다. 최근 대형 언어 모델(LLM)의 등장과 함께 few-shot 학습이 주목받고 있으며, 이는 데이터셋이 제한적인 환경에서 모델이 빠르게 학습할 수 있도록 돕는다. 본 연구에서는 Vision Transformer (ViT*를 특징 추출에 사용하고, Prototypical Networks를 통해 각 사람의 스타일을 학습하는 메타러닝 접근 방식을 제시한다. 세 명의 사람(A, B, C)으로부터 각각 3장의 학습용 샘플과 1장의 테스트용 샘플을 이용해 모델을 학습하고, 테스트 이미지에서 작성자를 예측하는 성능을 평가한다..

Ⅰ. 서 론

최근 대형 언어 모델(LLM)의 발전은 자연어 처리 분야에서 큰 영향을 미쳤으며, 특히 few-shot 학습의 중요성이 강조되고 있다[2]. LLM은 대규모 데이터셋 없이 적은 수의 예시로 새로운 작업을 학습할 수 있는 능력을 갖추고 있으며, 이는 모델이 빠르게 새로운 태스크를 학습하고 적용할 수 있게 해준다. 이러한 특성은 컴퓨터 비전 및 다양한 분야에서 메타러닝 기법을 효과적으로 적용하는 데 중요한 역할을 한다. 특히, few-shot 학습은 작은 데이터셋에서도 유효한 예측을 할 수 있는 모델을 구축할 수 있게 하여, 기존의 대규모 데이터셋 의존적인 방법에서 벗어나 효율적인 학습 전략을 가능하게 만든다.

본 연구는 메타러닝 기법을 활용하여 손글씨 작성자 식별 문제를 해결하고자 한다. Vision Transformer (ViT)를 사용하여 손글씨 이미지를 특징 벡터로 변환하고, Prototypical Network를 통해 각 사람의 글씨체 스타일을 대표하는 프로토타입을 학습한다. 이를 통해 제한된 데이터셋 환경에서 모델이 예측할 수 있는지 평가한다. Prototypical Network는 각 클래스에 대한 평균 특징 벡터(프로토타입)를 계산하고, 새로운 샘플이 주어졌을 때 이를 해당 프로토타입과 비교하여 예측을 수행한다.

실험은 학습용 샘플을 사용하여 모델을 훈련하고, 이후 테스트 샘플에 대해 예측을 진행한다. 예측이 정확하면 성능을 유지하고, 틀리면 정답을 제공하여 모델의 프로토타입을 업데이트한다. 이 과정은 모델이 학습과 피드백을 반복하며 성능을 지속적으로 향상시킬 수 있게 한다. 모델은 이러한 과정을 통해 점진적으로 성능을 개선하며, 결국 제한된 데이터셋에서도 효과적인 작성자 식별이 가능해진다. 본 연구는 ViT 기반의 메타러닝 모델이 제한된 데이터 환경에서 어떻게 성능을 향상시키고, 모델의 예측 능력을 지속적으로 개선하는지를 실험적으로 탐구한다.

Ⅱ. 본론

본 연구는 손글씨 작성자 식별을 위해 Vision Transformer (ViT)와 Prototypical Network를 결합한 모델을 사용한다[1][2]. ViT는 이미지를 여러 패치로 나누어 각 패치 간의 관계를 학습하고, 이를 통해 이미지의 전역적인 특성을 추출한다. 이 과정에서 생성된 특징 벡터는 Prototypical Network에 입력되어, 각 작성자의 스타일을 대표하는 프로토타입을 학습한다. 새로운 테스트 이미지가 들어오면, 해당 이미지의 특징 벡터와 각 프로토타입 간의 코사인 유사도를 계산하여 가장 유사한 작성자를 예측한다.

학습은 각 작성자에 대해 3개의 학습용 이미지와 1개의 테스트 이미지를 사용하는 방식으로 진행되었다. 먼저, 각 사람의 3개의 학습용 이미지를 사용하여 프로토타입을 생성한다. 이 프로토타입은 모델이 작성자 스타일을 인식하는 데 사용되는 기준 벡터로, .pt 파일로 저장된다. 이를 통해 모델은 각 사람의 고유한 스타일을 학습하고, 이 정보를 바탕으로 작성자 예측을 수행한다.

학습이 완료된 후, .pt 파일을 기반으로 모델을 테스트한다. 우선 a4.png라는 이름의 파일로 테스트를 진행했으며, 모델은 작성자 C를 예측했다.(그림1). 이후, 사용자가 정답 A를 제공하여 프로토타입을 갱신하였다. 같은 이미지를 사용한 두 번째 예측에서는 갱신된 프로토타입을 바탕으로 A를 정확히 예측했다(그림 2). 이 결과는 피드백 기반 학습이 점진적으로 성능 향상에 기여했는지를 보여준다.

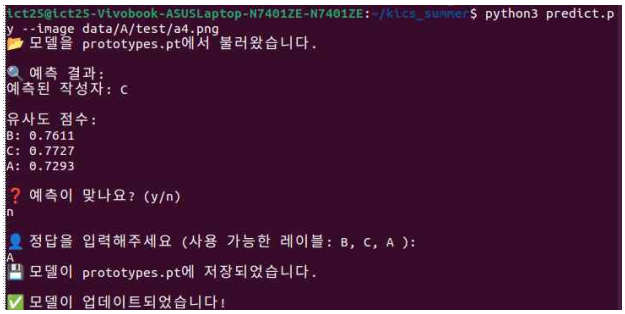

그림 1. 첫 번째 예측에서 잘못된 C를 예측한 결과.

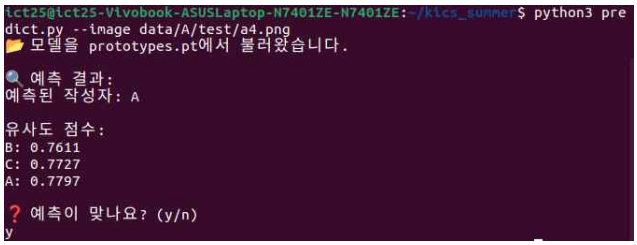

그림 2.피드백 후 정확히 A를 예측한 결과.

본 실험은 해당 모델을 사용하여 총 3명의 사람을 대상으로 각각 5장의 이미지를 이용해 총 15장의 테스트 이미지를 진행한다. 실험은 3단계로 나누어 진행하며, 각 단계마다 피드백 기반 학습을 통해 모델의 성능을 평가한다. 실험에 사용된 각 사람이 작성한 문구와, 볼펜의 굵기, 색깔 또한 각기 달라, 다양한 조건에서 모델의 손글씨 식별 성능을 테스트한다.
실험 결과, 첫 번째 단계에서는 모델이 15개의 예측 중 7개를 틀렸으며, 정확도는 53.33%로 나타났다. 이 과정에서 사용자는 틀린 예측에 대해 그때그때 정답을 제공하였고, 이를 바탕으로 프로토타입을 갱신하였다[3][4]. 이후 두 번째 단계에서는 모델의 예측 정확도가 향상되어, 15개의 예측 중 3개만 틀렸으며, 정확도는 80%로 증가하였다. 특히, 모델은 A와 B 작성자에 대해 정확한 예측을 보였고, C 작성자에 대해서는 여전히 몇 가지 오류가 발생하였다. 세 번째 단계에서는 모델이 모든 예측을 정확히 수행하였으며, 정확도는 93.33%에 도달하였다.

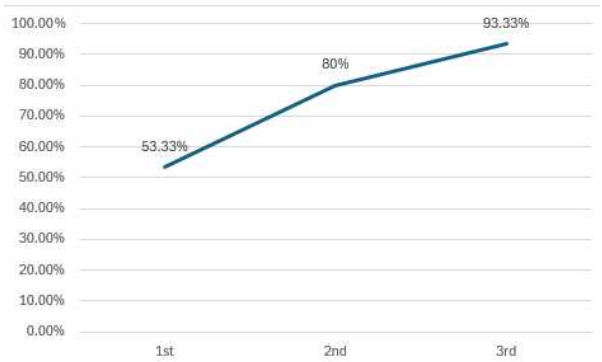

그림 3. 피드백 기반 학습 예측 정확도 향상

Ⅲ. 결론

본 논문에서는 소량의 필적 샘플만을 활용하여 작성자를 식별할 수 있는 Vision Transformer 기반의 메타러닝 모델을 제안한다. 제안된 모델은 피드백 기반 학습을 통해 초기 정확도 53.33%에서 최종 93.33%까지 성능이 크게 향상되었으며, 이를 통해 제한된 데이터 환경에서도 메타러닝 기법이 효과적으로 적용될 수 있음을 확인하였다.
특히, 본 연구는 대규모 데이터 확보가 어려운 국방이나 보안 등 특수 환경에서도 few-shot 학습을 통해 높은 정확도의 작성자 식별이 가능함을 보여준다[5].

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-대학ICT연구센터(ITRC)의 지원(IITP-2025-RS-2024-00437190, 50%)과 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원 - 학·석사연계ICT핵심인재양성 지원을 받아 수행된 연구임 (IITP-2025-RS-2022-00156394, 50%).

참 고 문 헌

[1] Koepf, M., Kleber, F., & Sablatnig, R., “Writer Identification and Writer Retrieval Using Vision Transformer for Forensic Documents,” *Document Analysis Systems*, pp. 352 - 366, Springer, 2022.

[2] Snell, J., Swersky, K., & Zemel, R., “Prototypical Networks for Few-shot Learning,” *Advances in Neural Information Processing Systems*, 2017.

[3] He, S., & Schomaker, L., “Deep Adaptive Learning for Writer Identification based on Single Handwritten Word Images,” *arXiv preprint*, arXiv:1809.10954, 2018.

[4] Srivastava, A., Chanda, S., & Pal, U., “Exploiting Multi-Scale Fusion, Spatial Attention and Patch Interaction Techniques for Text-Independent Writer Identification,” *arXiv preprint*, arXiv:2111.10605, 2021.

[5] Nguyen, H. T., Nguyen, C. T., Ino, T., Indurkhya, B., & Nakagawa, M., “Text-independent writer identification using convolutional neural network,” *arXiv preprint*, arXiv:2009.04877, 2020.