한국컴퓨터정보학회지
제28권 제2호, 2020. 12.

딥러닝 모델에서의 흐름정보 기반 지식증류 기법 분석

• 배지훈 (대구가톨릭대학교)

Ⅰ. Introduction

현재, 전 세계적으로 4차 산업혁명 시대가 도래하면서 이에 따른 기술 발전으로 인하여 산업구조가 빠르게 변화하고 있다. 특히, 4차 산업혁명의 핵심기술인 인공지능을 기존 산업/기술 분야뿐만 아니라 사회 · 경영 · 문화 · 예술 등의 다양한 분야에 적용하려는 움직임이 눈에 띄게 드러나고 있으며, 이러한 변화의 핵심에는 인간의 뇌를 모방한 딥러닝(Deep-learning) 기술이 있다. 최근, 딥러닝 기술은 각종 분야에서 기계학습 최고의 성능을 갱신하고 있으며, 이러한 추세에 따라 구글, 페이스북, 아마존 등 세계 유수 기업뿐만 아니라, 네이버, 카카오, 삼성전자, SK텔레콤 등 국내 여러 기업에서도 딥러닝 기술을 적극 개발, 활용하고 있다.

딥러닝 학습은 그 최고의 성능을 달성하기 위하여 일반적으로 구조가 복잡한 심층 신경망(deep neural network, DNN) 네트워크 형태를 가지고 있고, 그에 따른 많은 학습 데이터들을 필요로 한다. 특히, 심층 신경망에서 오늘날의 딥러닝 기술 분야의 발전을 가져다 준 합성곱 신경망(Convolutional neural networks: CNN)은 보통 수많은 학습변수 및 매개변수들을 가지고 있으며, 네트워크의 표현력을 증가시키기 위하여 모델 구조가 매우 복잡하고 많은 층(layer)들로 이루어져 있다[1-5]. 일반적으로, 이러한 CNN 모델로부터 최적의 학습 결과를 도출하기 위해서는, 라벨링(labeling)이 있는 많은 양의 학습데이터 확보 및 많은 시간과 노력, 고비용의 컴퓨팅 하드웨어 자원들을 필요로 한다. 하지만, 이미지 인식 대회인 ILSVRC(ImageNet Large Scale Visual Recognition Challenge)[6]에서 우승한 고성능의 딥러닝 모델들을 처음부터 새로이 학습하는 것이 아니라 다양한 오픈소스 소프트웨어 플랫폼들을 통하여 사전학습 완료된 고성능 모델들을 쉽게 다운 받아 활용할 수 있다. 즉, 구조가 복잡하고 고사양의 DNN 모델을 처음부터 학습하는 것이 아니라 사전에 최적으로 학습 완료된 모델 구조를 그대로 재사용하는 전이학습(Transfer learning) 기법을 활용하는 것이다[7].

전이학습은 딥러닝 학습의 필수 요소인 데이터 부족 문제를 극복할 수 있는 학습기법으로, 주요 특징들이 자동으로 추출되어 특징맵(feature map)에 임베딩되어 있는, 분류기를 제외한 CNN 모델 구조 대부분을 직접 재사용하기 때문에, 적은 양의 학습데이터를 이용하더라도 유사 도메인에 대하여 최적의 추론 성능을 도출할 수 있는 장점을 가지고 있다. 하지만, 기존 전이학습은 사전 학습 완료된 모델에 의존적으로 재학습되기 때문에, 컴퓨팅 자원이 제한되는 분야에 적용하는데 그 한계점이 발생할 수 있다.

이러한 문제점을 극복하기 위하여, 사전학습 완료된 DNN 모델에서 유용한 지식을 추출하고, 이를 다른 DNN 모델로 이전하여 학습을 수행하는 지식증류(knowledge distillation) 기법이 대두되었다. 처음 소개된 지식증류 기법은 소프트맥스 층에서 출력된 확률분포 정보를 완화하여 다른 대상 모델로 이전하는 방식이다[8]. 본 논문에서는 기존 방식과 달리 복잡한 DNN 모델 내부의 지식을 표현하는 흐름정보(flow information)[9]를 추출하여 다른 DNN 모델로 이전하여 학습을 수행하는 다양한 흐름정보 기반 지식증류 기법들에 대하여 살펴보고자 한다.

Ⅱ. 흐름정보 기반 지식증류

흐름정보는 두 개의 층 사이에 출력되는 해당 특징맵들이 변화하는 흐름을 정의한 것으로, 수학적으로 다음의 두 특징

맵 사이의 상관관계를 나타내는 그람행렬 G (Gram matrix)로 표현된다[9].

$$G = \left[g_{ij}(x;W) \right]_{i=1,2,\ldots,N,\, j=1,2,\ldots,M,} \quad (1)$$

여기서, x는 입력 이미지를, W는 가중치를 각각 나타내고, $g_{ij}(x;W)$는 다음 식과 같이 정의된다.

$$g_{ij}(x;W) = \sum_{k=1}^{K} \sum_{l=1}^{L} \frac{F^{u_{s,t,i}}(x;W) \times F^{l_{s,t,j}}(x;W)}{K \times L} \quad (2)$$

여기서, K와 L은 특징맵의 높이와 폭을 의미하고, F^u와 F^l은 각각 상위층에서의 특징맵과 하위층에서의 특징맵을 각각 나타낸다.

흐름정보 지식증류는 그림 1과 같이 DNN의 한 종류인 사전학습 완료된 residual network(ResNet)[4] 구조에서 낮은 층에서부터 높은 층으로 복수 개의 특징맵 상관관계 흐름정보들을 추출하고, 상기 추출된 정보들을 동종의 다른 ResNet 모델로 이전하는 단계별 학습을 수행한다. 즉, 단계 1에서는 추출된 흐름정보를 다른 모델로 이전하는 지식전이(knowledge transfer) 학습을 수행하고, 단계 2에서는 단계 1에서 학습한 모델의 가중치를 초기 가중치로 설정하여 확률분포를 나타내는 소프트맥스(softmax) 층을 활용한 전통적인 지도학습(supervised learning)을 수행한다.

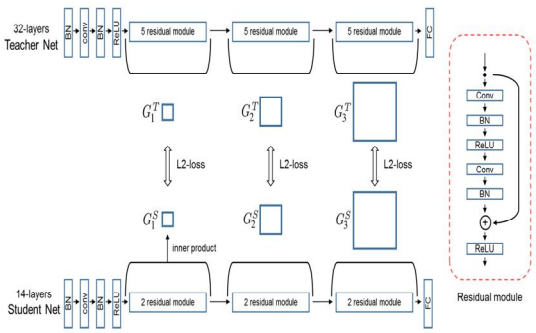

Fig. 1. 흐름정보 지식증류를 이용한 DNN 모델 전체 구성도[9]

DNN 모델의 내부 정보를 표현한 흐름정보 지식증류를 이용하게 되면 기존 특징맵 출력 정보만을 이용하는 힌트정보 지식증류 기법[10]보다 더 높은 이미지 분류 정확도를 제공하는 것을 실험적으로 관찰할 수 있다. 기존에는 단계 1에서 복수 개의 흐름정보 전달 시, 동일한 가중치를 설정하여 지식전이를 수행하였으나, 계층별 중요도에 따라 흐름정보에 적용되는 가중치를 적응적으로 분배하는 분석 연구를 수행하면 더 높은 정확도를 가지는 학습결과가 기대된다.

III. 밀집흐름정보 기반 지식증류

밀집흐름정보 지식증류 기법[11]은 지식증류에 있어서 II장의 흐름정보 추출을 강화한 것으로, 그림 2와 같이 특징맵 상관관계 흐름정보들을 서로 중첩하여 밀집하게 추출하는 방식으로, 학습 방식은 다음의 두 단계로 구성된다.

- 단계 1: 밀집하게 추출된 흐름정보들을 순차적으로 다른 DNN 모델로 이전하여 학습을 수행
- 단계 2: 단계 1에서 학습한 DNN 모델의 가중치를 초기 가중치로 설정하여 소프트맥스층을 활용한 전통적인 지도학습을 수행

단계 1에서, 기존 흐름정보 학습기법은 추출된 복수 개의 흐름정보들에 대한 동시 이전 학습을 수행하는 반면, 밀집흐름정보 학습기법에서는 다른 딥러닝 모델로 정보이전 시, 추출된 밀집흐름 정보들을 단계별로 나누어 순차적으로 이전학습을 수행하였다.

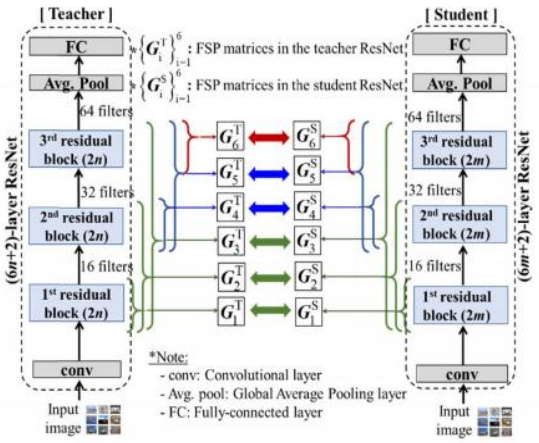

Fig. 2. 밀집흐름정보 지식증류를 이용한 DNN 모델 전체 구성도[11]

밀집흐름정보 지식증류는 저수준(low-level) 특징맵들이 추상적인 고수준(high-level) 특징맵들로 변화하는 상관관계 흐름정보들을 서로 중첩되고 밀집하게 추출함으로써, 사전학

습 완료된 딥러닝 모델의 내부 지식추출에 대한 표현력을 좀 더 확장하고 강화한 기법으로 해석할 수 있다. 따라서, 실험 결과에서도 기존 흐름정보 지식증류 기법[9]보다 더 우수한 지식전이 성능을 보여줌을 관찰할 수 있다.

IV. 흐름정보 지식증류에 대한 최적화 기법

II 장에서 소개한 흐름정보 지식증류 기법에서 추출된 흐름정보를 다른 학습대상 DNN 모델로 이전할 경우, 학습에 필요한 관련 파라미터들에 대한 전통적인 그리드 검색(grid search) 방법을 이용하여 학습을 수행하였다. 하지만, 이러한 기존 방식은 네트워크 구조가 복잡한 DNN 모델에 대하여 일반적으로 학습 시간이 오래 소요되고, 여러 번의 학습시도에도 최적의 해에 도달하기 어려울 수 있다. 이러한 한계점을 극복하기 위하여, [12]에서는 그림 3과 같이 흐름정보 지식증류 기법에 필요한 중요한 학습 파라미터들에 대하여 베이지안 최적화(Baysian Optimum) 기법을 적용하여 사전 학습 완료된 DNN 모델로부터 흐름정보를 추출하고 빠른 학습이 가능한 알고리즘을 제안하였다.

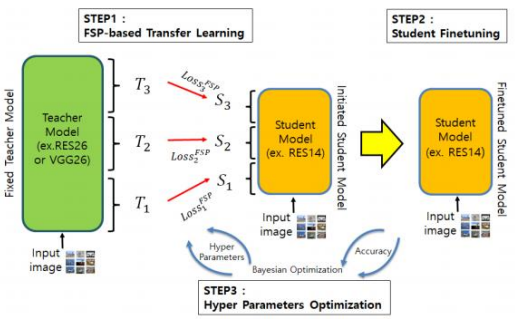

Fig. 3. 흐름정보 지식증류의 학습 파라미터 최적화 구성도[12]

[12]에서 제안한 기법은 사전학습 완료된 DNN 모델과 대상 DNN 모델 모두 그 구조가 동일한 동종 모델들(ResNet[4]-ResNet[4]) 혹은 서로 다른 이종 모델(ResNet[4]-VGGNet[3])들에 대하여 각각 비교 실험들을 수행하였으며, 흐름정보 기반 지식증류에 결합한 기존 그리드 검색 기법 대비 학습속도가 더 빠르면서 0.1%~0.3%의 높은 정확도를 보여주었다. 이러한 최적화 알고리즘은 기존 흐름정보 지식증류 기법보다 더 복잡한 III 장에서 소개한 밀집흐름정보 지식증류에도 확장하여 보다 더 높은 정확도를 가지는 학습모델 생성에 적용할 수 있을 것으로 기대된다.

V. 적대적 생성 신경망 기반의 흐름정보 지식증류

앞에서 기술하였던 흐름정보 지식증류 기반 학습기법은 유클리디안(Euclidean) 거리가 최소가 되도록 비용함수(cost function)를 설정하여 학습을 수행하였다. 이와 달리, 적대적 생성 신경망(generative adversarial network, GAN)[13] 구조를 적용하여 추출된 흐름정보를 다른 대상 모델로 이전하여 학습을 수행하는 새로운 연구가 수행되었다[14]. 이는 생성자(generator)를 학습 대상DNN 모델로 설정하고, 상기 DNN 모델에서 추출한 흐름정보들을 MLP(multi-layer perceptron)로 구성된 판별자(discriminator)가 사전 학습 완료된 DNN 모델로부터 추출된 흐름정보들과 서로 비교하여 구분하지 못할 때까지 적대적 학습을 수행하는 방식이다. 이러한, 적대적 생성 신경망과 결합한 흐름정보 지식증류에 대한 학습을 수행하게 되면, 그림 4와 같이 학습 대상 DNN 모델이 사전 학습 완료된 DNN 모델로부터 추출된 흐름정보에 대한 특징 분포를 보다 정확하게 닮아가도록 학습할 수 있음을 실험적으로 관찰할 수 있다.

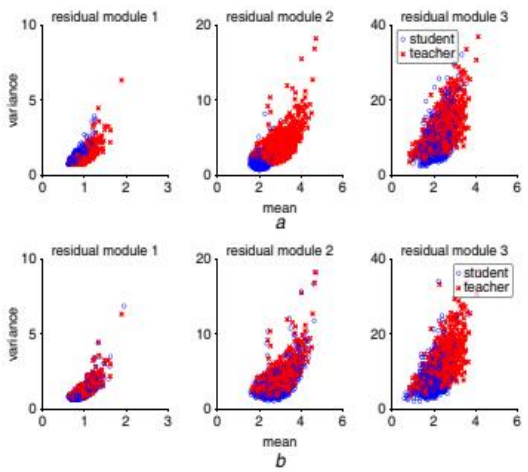

Fig. 4. 흐름정보에 대한 특징 분포 결과[14]. (a)기존 유클리디안 학습 방식. (b) 적대적 생성 신경망 기반 학습 방식

이 기법 또한 구조가 복잡한 밀집흐름정보 지식증류에 적대적 생성 신경망을 적용하여, 중첩되고 밀집한 흐름정보를 이전받는 학습 대상 DNN 모델이 사전학습 완료된 DNN 모델의 밀집흐름정보 분포와 거의 일치되게 학습하여 그 정확도 성능을 향상시킬 수 있다.

Ⅳ. 지식증류 주요 정보 재구성

일반적으로, 기존 흐름정보 기반 지식증류 기법은 계층별 특징맵 결과들을 모두 이용하여 흐름정보에 대한 상관관계를 계산하였다. 하지만, 구조가 복잡한 DNN 모델의 경우, 일반적으로 층이 높아질수록 특징맵 개수가 증가되는 구조를 가지고 있으며, 계층별로 출력되는 수많은 특징맵들에 불필요한 redundancy가 존재할 수 있다. 또한, 기존 지식증류 방식은 두 층 사이에 동일한 크기의 공간적 분포를 가지는 특징맵들을 필요로 하기 때문에, 지식증류를 수행하고자 하는 모델 구조에 따라 제한점이 발생할 수 있다.

따라서, 이러한 문제점을 해결하기 위하여 [15]에서는 그림 5와 같이 SVD(singular value decomposition) 기법을 이용하여 서로 다른 두 특징맵들을 각각 분해하여 중요한 정보가 포함된 특징맵들만 추출하고, 가우시안 함수(Gaussian function)를 통한 두 특징맵 사이의 관계를 유도한 일반화된 흐름정보 기반 지식증류 기법을 제안하였다. 이 기법에 따르면, 지식증류를 수행하고자 하는 구조적 제한 없이 중요 정보를 포함한 특징맵들을 추출하여 지식증류를 수행할 수 있으며, 학습 대상 모델에 대한 지식전이 학습 수행 시, 기존 흐름정보 지식증류 기법 대비 정확도 성능이 향상된 것을 관찰할 수 있다.

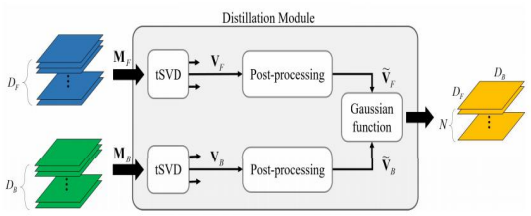

Fig. 5. 주요 정보를 추출하여 재구성하는 지식증류 구성도[15]

Ⅲ. Conclusions

본 논문에서는 딥러닝 모델에서의 지식증류 방법으로써, 흐름정보를 기반으로 하는 여러 가지 기법들에 대하여 살펴보았다. 흐름정보 기반 지식증류는 심층신경망 내부의 지식 표현을 모델링한 기법으로, 기존 전이학습 방법의 장점인 모델 재사용과 함께 유사 도메인에 학습대상 모델을 적응적으로 선택하여 지식이전을 수행할 수 있는 장점을 가지고 있다. 또한, 사전 학습 완료된 지식제공 없이 학습하는 모델보다 더 높은 정확도를 달성할 수 있을 뿐만 아니라, 더 빠른 최적화 학습도 가능함을 실험적으로 보여주었다. 향후, 이러한 지식증류 기법과 결합하여 딥러닝 모델 학습 파라미터에 대한 자동 최적화 혹은 모델생성 자동화 연구들로 발전될 것으로 보인다.

ACKNOWLEDGEMENT

이 결과물은 2020년도 대구가톨릭대학교 교내연구비 지원에 의한 것임

REFERENCES

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convoluti onal neural networks," in 26th Annual Conf. Neural Information Process. Sys. (NIPS) 2012, Stateline, Nevada, USA, Dec. 3-8, 2012, pp. 1106-1114.

[2] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in Proc. 2015 IEEE Conf. Comput. Vision Pattern Recogn. (CVPR), Boston, USA, June 7-12, 2015, pp. 1-9.

[3] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in Proc. 5th Int. Conf. Learning Represent. (ICLR), San Diego, USA, May 7-9,

2015, pp. 1-14.
[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vision Pattern Recogn (CVPR), Las Vegas, USA, Jun. 26-Jul. 1, 2016, pp. 1-12.
[5] G. Huang, Z. Liu, L. V. D. Maaten, and K. Weinberger, "Densely connected convolutional networks," in Proc. 2017 IEEE Conf. Comput. Vision Pattern Recogn. (CVPR), Honolulu, USA, Jul. 21-26, 2017, pp. 2261-2269.
[6] ImageNet, Large Scale Visual Recognition Challenge (ILSVRC), http://www.image-net.org/challenges/LSVRC/
[7] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," Int. J. Comput. Vis. (IJCV), vol. 115, no. 3, pp. 211-252, 2015.
[8] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," arXiv preprint arXiv:1503.02531, pp. 1-19, 2015.
[9] J. Yim, D. Joo, J.-H. Bae, and J. Kim, "A gift from knowledge distillation: Fast optimization, network minimization, and transfer learning," in Proc. of 2017 IEEE Conf. Comput. Vision Pattern Recogn. (CVPR), Honolulu, USA, Jul. 21-26, 2017, pp. 7130-7138.
[10] A. Romero, N. Ballas, S. E. Kahou, A. Chassang, C. Gatta, and Y. Bengio, "Fitnets: Hints for thin deep nets," in Proc. 5th Int. Conf. Learning Represent. (ICLR), San Diego, USA, May 7-9, 2015, pp. 1-13.
[11] J.-H. Bae, D. Yeo, J. Yim, N.-S. Kim, C.-S. Pyo, and J. Kim, "Densely distilled flow-based knowledge transfer in teacher-student framework for image classification," IEEE Transactions on Image Processing, vol. 29, pp. 5698-5710, 2020.
[12] K. Kim and J.-H. Bae, "Important parameter optimized flow-based transfer learning technique supporting heterogeneous teacher network based on deep learning," Journal of KIIT, vol. 18, No. 3, pp. 21-29, 2020.
[13] Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al. "Generative adversarial nets", Advances in Neural Information Processing Systems, Canada, December 2014, pp. 2672-2680.
[14] D. Yeo and J.-H. Bae, "Multiple flow-based knowledge transfer via adversarial networks," Electronics Letters, Vol. 551, No. 18, pp.989-992, Sept. 2019.
[15] S. Lee and B.-C. Song, "Knowledge transfer via decomposing essential information in convolutional neural networks," IEEE Transactions on Neural Networks and Learning Systems, pp.1-12, 2020.

저 자 소 개

Ji-Hoon Bae received the B.S. degree in electronic engineering from Kyungpook National University, Daegu, Korea, in 2000, and the M.S. and Ph.D. degrees in Electrical Engineering from Pohang University of Science and Technology (POSTECH), Pohang, Gyeongsangbuk-do, Korea, in 2002 and 2016, respectively From 2002 to 2019, Dr. Bae was with Electronics and Telecommunications Research Institute (ETRI), Daejeon, Korea, as a principal researcher. He joined the faculty of the Department of AI・Big Data Engineering, Daegu Catholic University, Gyeongsan-si, Gyeongbuk, Korea in 2019. He is interested in deep learning and transfer learning, radar imaging, and optimized techniques.