

감정 간의 관계를 고려한 지도 대조 학습 기반 감정 인식

유동제¹, 전경훈², 최주환³, *김영빈⁴

¹중앙대학교 컴퓨터공학과 ²중앙대학교 수학과

³중앙대학교 AI학과 ⁴중앙대학교 첨단영상대학원

e-mail: pass120@cau.ac.kr, kyunghoon@cau.ac.kr,

gold5230@cau.ac.kr, ybkim85@cau.ac.kr

Emotion Relationship-based Sentiment Analysis via Supervised Contrastive Learning

Dongje Yoo¹, Kyunghoon Jeon², Juhwan Choi³, and *Youngbin Kim⁴

¹School of Computer Science and Engineering, Chung-Ang University

²Department of Mathematics, Chung-Ang University

³Department of Artificial Intelligence, Chung-Ang University

⁴Graduate School of Advanced Imaging Science, Multimedia & Film, Chung-Ang University

Abstract

Emotion recognition has been researched in many fields aiming to highly match emotion labels beyond all kinds of data. Recent studies have suggested utilizing the 2d arousal-valence model for emotion-related tasks. Inspired by the previous works, we propose a sentiment analysis technique considering the similarity between emotion labels. Our technique is made up of Supervised Contrastive Learning Loss grafting together with Curriculum Learning. From the proposed technique, we observed a further performance gain on emotion recognition, and we achieved decent work for three of the public datasets that we used for experiments.

I. 서론

감성 분석은 텍스트 데이터에서 표현된 감정과

의견을 분석하는 과정이다. 인간은 다양한 매체를 통해 자신의 생각과 감정을 표현하고 있기에, 감성 분석은 모든 유형의 데이터에서 감성을 이해하는 필수적인 도구가 되어가고 있다.

감성 분석은 딥러닝 기반 자연어처리 기술이 발전함에 따라 CNN [1], LSTM [2] 등의 신경망 모델에서 BERT [3] 등의 사전 학습 언어 모델을 적용하는 방향으로 이어져 왔다. 대규모의 데이터에 대해 사전 학습된 언어 모델을 미세 조정하며 감성 분류 작업을 학습하는 과정에서 감성 라벨은 원-핫 인코딩으로 구성되어 교차 엔트로피 함수를 통해 학습된다. 그러나 이러한 원-핫 인코딩을 활용한 분류는 라벨 간의 상관관계를 고려하지 않는 문제점이 있다. 예를 들어 Joy와 Surprise는 Sad와 비교해 더욱 고양된 상태라는 공통점이 있지만, 원-핫 인코딩은 이런 라벨 간 유사성을 고려하지 않는다.

본 연구는 우리는 이러한 문제를 보완하기 위해 Contrastive Learning 기법을 감성 인식에 접목한다. Face-Net [4]에서 처음 제안된 Triplet Loss는 동일한 라벨을 가지는 데이터는 특징 공간 상에서의 거리를 가깝게 하고, 다른 라벨을 가지는 데이터는 멀어지게 하는 학습 방법이다. 이는 서로 구분하기 어려운 데이터에 대해

Anchor와의 거리를 조정하여 분류를 용이하게 만드는 장점이 있다. 그러나 이 방식은 Triplet Loss를 계산할 때 한 번에 3개의 데이터 밖에 비교하지 못하는 단점이 있다. 이러한 단점을 개선하고자 N-Pair Loss [5]가 제안되었다. 이는 비지도 학습 시 배치 내에서 다른 라벨값을 가지는 Negative pair의 모든 데이터와의 거리를 멀어지게 하며, 같은 라벨을 갖는 Positive pair의 거리는 가깝게 하는 방법이다. 이를 응용하여 여러 클래스로 구성된 지도학습에 적용하고자 고안된 방식이 Supervised Contrastive Learning (SCL) [6]이다.

하지만 감정 인식에서는 SCL Loss의 Negative pair의 구성에 있어서 기존과 다른 접근법을 가져야 한다. 감정인식 작업에 있어 라벨들 간의 유사성과 그에 따른 분류의 난이도 차이 또한 고려해야 하기 때문이다. 그러나 기존 SCL Loss를 활용한 감정인식은 이러한 라벨 간의 유사성을 고려하지 않고 Anchor와 다른 모든 라벨을 Negative pair로 활용한다. 그렇기에 라벨의 유사도를 고려하여 SCL Loss를 재구성할 필요가 있다.

한편, 각성도(arousal)와 감정가(valence)를 축으로 한 2차원 감정 분류 모델인 2d arousal-valence model이 1980년에 처음 제안되었다[7]. 이후 이 모델은 자연어 처리 뿐만 아닌 음성인식 등의 다양한 분야의 감정 인식 작업에서 활용되고 있다[8].

본 연구에서는 2d arousal-valence model을 활용해 감정 라벨 간의 유사도를 측정한다. 해당 모델을 나타낸 그림 1에서 θ 는 두 라벨 사이의 유사도를 의미한다. 또한, 이렇게 정의된 유사도를 활용해 쉬운 작업부터 어려운 작업까지 점진적으로 학습하는 커리큘럼 학습을 진행한다. 기존 연구는 두 라벨 사이의 거리가 가까울수록 Negative로 분류하기 어려운 것임을 고려하여 커리큘럼 학습을 진행하였으며, 그 결과 성능 향상을 거두었다[9]. 이를 통해 그림 1 모델에서 두 라벨 간에 가깝게 위치할수록 모델이 구분하기 더욱 어려운, 유사한 라벨임을 알 수 있다.

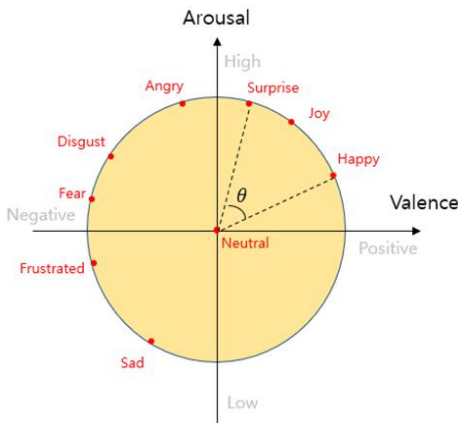


그림 1. 2d arousal-valence model

본 논문에서는 기존 Supervised Contrastive Learning Loss에 라벨 간의 유사도와 난이도의 개념을 도입하여 Negative pair 및 Positive pair를 변형하며, 나아가 이를 커리큘럼 학습과 접목하여 유사한 라벨의 분리를 효과적으로 수행한다. 실험을 통해 제안된 방식을 접목하였을 때 감정 라벨 인식의 정답률이 세 개의 데이터셋에서 정확도와 F1 score가 평균 0.75%p, 1.11%p 상승한 것을 확인할 수 있었다.

II. 본론

본 장에서는 2d arousal-valence model을 고려한 감정인식의 Supervised Contrastive Learning 방법과 이를 커리큘럼 학습과 결합한 방법을 소개한다.

2.1 Supervised Contrastive Learning

Supervised Contrastive Learning은 Anchor가 되는 데이터와 배치 내에 있는 모든 동일한 라벨의 데이터를 Positive pair로 여기고, 나머지 다른 라벨의 데이터는 Negative pair로 여긴다. 이를 통해 Anchor와 Positive pair와는 거리가 가깝게 하고, Negative pair와는 거리가 멀어지도록 학습한다. 이렇게 배치 내에 모든 데이터와의 거리를 비교하며 손실함수를 구성함의 목적은, 같은 쌍의 데이터를 참조하여 특정 공간 상에서 데이터 간의 거리를 원활하게 조절하는데 있다.

BERT 모델을 통해 N개의 문장으로 구성된 배치를 인코딩한 벡터의 집합은 $I = [z_1, z_2, \dots, z_N]$ 와 같이 나타낼 수 있다. 이렇게 얻은 벡터 간의 코사인 유사도를 나타내는 $G(z_i, z_j) = \cos \theta$ 는 두 벡터를 서로 내적하여 얻을 수 있는 스코어 함수이다. 여기에 정규화를 위해 지수 승을 적용한 $F(z_i, z_j) = \exp\left(\frac{G(z_i, z_j)}{T}\right) = \exp\left(\frac{\cos \theta}{T}\right)$ 를 Positive pair와 Negative pair의 계산에 활용한다. θ 는 두 벡터 사이의 각도이며, T 는 입력의 세기를 조절하는 초매개변수이다.

이때 Positive pair와 Negative pair에 대한 식은 다음과 같다.

$$P_{sup}(i) = \sum_{z_p \in P(i)} F(z_i, z_p), N_{sup}(i) = \sum_{z_j \in A(i)} F(z_i, z_j)$$

이로써 손실함수는 다음과 같다.

$$L_i^{sup} = -\log \frac{1}{|P(i)|} \frac{P_{sup}(i)}{N_{sup}(i)}$$

$A(i) \equiv I \setminus \{z_i\}$ 는 배치 내에서 자신을 제외한 모든 벡터들의 집합이며, $P(i)$ 는 배치 내에서 z_i 와 동일한 라벨을 가지는 벡터들의 집합이다.

2.2 Negative Pair 가중치 구성

2d arousal-valence model에서 인접한 라벨을 갖는 데이터간의 유사도는 멀리 떨어진 라벨을 갖는 데이터간의 유사도보다 더욱 높게 나타나고, 특징 공간 상에서도 서로 가깝게 위치한다. 이렇게 데이터간의 거리가 상대적으로 가까워서 다른 Negative에 비해 Negative로 분류되기 어려운 관계를 Hard-negative라고 한다. Hard-negative 관계에서 모델이 두 데이터를 용이하게 구분할 수 있도록 특징 공간 상에서의 거리가 멀어지도록 유도하기 위해, 기존 SCL Loss의 Negative pair 손실 함수에 아래와 같이 가중치를 부여한다.

$$F_{neg}(z_i, z_j) = \exp\left(\frac{\cos \theta}{T}\right) * \exp\left(\frac{\cos \alpha}{T}\right)$$

$$\cos \alpha = \begin{cases} z_i \cdot z_j, & z_i \cdot z_j > 0 \\ 0, & z_i \cdot z_j \leq 0 \end{cases}$$

위 수식에서 α 는 2d arousal-valence model에서 두 라벨 사이의 유사도이다. 90도 이상 차이가 나면 0의 값을 가지며, 90도 이내이면 인접한 라벨로 간주한다. F_{neg} 는 다음과 같이 나타낼 수 있다.

$$F_{neg}(z_i, z_j) = \exp\left(\frac{\cos \theta + \cos \alpha}{T}\right)$$

Hard-negative에 해당할수록 Anchor와의 거리가 가까우면 더욱 Negative로 분류되기 어렵다. 이러한 Hard-negative를 Negative로 올바르게 분류하기 위해서는 Hard-negative와 Anchor 간의 특징 공간 상의 거리가 더욱 멀어져야 한다. 이를 위해 아래 수식과 같이 Hard-negative에 해당할수록 손실 함수 값이 커지도록 구성했다.

$$N_{neg}(i) = \sum_{z_j \in A(i)} \exp\left(\frac{\cos \theta + \cos \alpha}{T}\right)$$

2.3 Positive Pair 가중치 구성

소프트 라벨은 학습 시 정답이 아닌 클래스도 일부 정답으로 인정하는 학습 방법으로, 과적합을 방지하고 표상의 일반화를 더욱 용이하게 하는 효과가 있다[10].

본 연구에서는 이러한 소프트 라벨의 개념을 활용하여, 정답은 아니지만 정답과 유사한 라벨에 대해서 기존 SCL Loss의 Positive pair로 일부 인정하는 변형된 학습법을 제안한다. 이때, 2d arousal-valence model에서의 라벨 간 유사도를 Positive pair에서 일부 정답으로 인정하는 기준으로 활용하였다.

$$S_{pos}(i) = \sum_{z_j \in S(i)} \exp\left(\frac{\cos \theta}{T}\right) * \cos \alpha$$

위 수식과 같이 벡터의 유사도에 $\cos \alpha$ 를 가중치로

곱한 만큼만 Positive pair로 인정한다. 이로 인한 최종 손실 함수는 다음과 같다.

$$L_i^{pos} = -\log \frac{1}{|P(i)|} \frac{P_{sup}(i)}{N_{sup}(i)} - \log \frac{1}{|N_{S(i)}|} \frac{S_{pos}(i)}{N_{sup}(i)}$$

where $N_{S(i)} \geq |S(i)|$

$S(i)$ 는 i 와 코사인 유사도가 $\pi/2$ 이내에 있으면서 동일한 클래스는 아닌 벡터들의 집합이다. $N_{S(i)}$ 는 Positive pair의 비율을 조정하기 위한 조매개변수이다.

변형된 손실함수의 분자 부분은 Anchor와의 거리를 가깝게 해야 하는 쌍이지만, 최종적으로는 Hard-negative와의 공간 상의 거리를 멀어지게 해야 한다. Positive pair 가중치 구성을 단독으로만 사용하면 Hard-negative와의 거리를 가까이 함에 따라 손실함수 값이 작아지기에 오히려 Hard-negative와의 거리를 가까이 하는 역효과가 날 수 있다. 따라서 이 방법은 단독으로 사용하지 않고 아래의 커리큘럼 학습과의 결합을 통해서만 사용하였다.

2.4 커리큘럼 학습과의 결합

커리큘럼 학습은 난이도를 측정하여 쉬운 작업을 먼저 학습하고, 어려운 작업을 이후에 학습하는 방식으로 진행된다. 이에 따라 난이도를 측정하는 과정은 커리큘럼 학습에서 핵심적인 역할을 수행한다. 이 개념을 감정인식에 적용해보면, 2d arousal-valence model에 따라 유사도가 높은 라벨들은 Hard-negative이기에 이들은 다른 Negative에 비해 상대적으로 모델이 올바르게 분류하기 어렵다. 이와 같이 작업 난이도의 어려움을 정의할 수 있기에 감정 인식 작업에서 커리큘럼 학습의 효과를 기대할 수 있다.

2.4.1 Positive Pair에 적용

소프트라벨의 개념을 활용하여 Hard-negative를 Positive pair에 추가한 것은, 모델이 다른 Negative 보다 Hard-negative를 Positive로 잘못 판단하기 쉽다는 이유에 기인한다. 반대로 Hard-negative는 다른 Negative에 비해 상대적으로 올바르게 분류되기 어렵다. 이러한 문제를 해결하기 위해 기존 Positive pair의 수식에 커리큘럼 학습을 적용해 다음과 같은 식으로 나타낼 수 있다.

$$S'_{pos}(i) = \sum_{z_j \in S(i)} \exp\left(\frac{\cos \theta}{T}\right) * \cos \alpha * \gamma$$

$$L_i^{pos} = -\log \frac{1}{|P(i)|} \frac{P_{sup}(i)}{N_{sup}(i)} - \log \frac{1}{|N_{S(i)}|} \frac{S'_{pos}(i)}{N_{sup}(i)}$$

γ 는 커리큘럼 비율로, 학습이 진행됨에 따라 1에서

0으로 감소한다. 이로써 Hard-negative를 정답으로 인정하는 비율이 점점 적어지므로 학습 난이도는 점점 어려워진다.

2.4.2 Negative Pair에 적용

앞선 방법에서는 모델에게 라벨의 유사도가 클 수록 손실 함수값을 크게 줌으로써 더욱 Hard-negative를 구분하게 만들었다. 본 장에서는 이를 확장해 커리큘럼 학습에 적용한다. 라벨 유사도를 고려해 가중치를 준 L_i^{sup} 에서 학습이 진행될 수록 Vanilla L_i^{sup} 가 되도록 손실 함수를 아래와 같이 구성하였다.

$$N'_{neg}(i) = \sum_{z_i \in A(i)} \exp\left(\frac{\cos \theta + \gamma \cdot \cos \alpha}{T}\right)$$

$$L_i^{sup} = -\log \frac{1}{|P(i)|} \frac{P_{sup}(i)}{N'_{neg}(i)}$$

γ 는 위와 마찬가지로 학습이 진행됨에 따라 1에서 0으로 감소하면서 유사한 라벨로 인해 증가하는 손실 함수의 값 또한 점점 감소한다.

III. 실험

이 장에서는 본 논문에서 제안하는 기법의 효과를 검증하기 위한 실험 구성과 실험 결과에 대해 설명한다.

3.1 실험 환경

사전학습 모델인 BERT를 변형된 Supervised Contrastive Learning을 통해 학습 시킨 후, BERT의 매개변수를 고정시킨 뒤 미세조정을 진행하였다. 커리큘럼 학습은 모든 배치를 독립적으로 학습하는 One-step 전략을 사용했고 매 Iteration마다 커리큘럼 비율을 감소시키며 학습했다. 배치의 크기와 문장의 최대 길이는 모든 학습에서 128, 64로 통일했다. 변형된 Supervised Contrastive Learning, 미세조정의 학습률은 $1e-6$ 로 하였다. 변형된 Supervised Contrastive Learning에 대해서는 8 에폭, 미세조정에 대해서는 16 에폭 동안의 Patience를 두고 Weighted-F1 score를 기준으로 Early Stopping을 적용하였다. 실험에 사용된 GPU는 GTX 1080Ti 1대이다. $N_{s(i)}$ 는 $|S(i)|^2$ 로 설정했다. 실험에는 총 세가지 데이터셋을 사용하였다.

Emotions [11] 데이터셋은 Angry, Fear, Joy, Sad, Surprise로 구성된 트위터 영문 메세지 데이터셋이다. 본래 구성에는 love 라벨도 있지만 이는 2d arousal-

valance model에 포함되지 않기에 임의로 삭제하였다. Friends¹ 데이터셋은 Neutral, Joy, Surprise, Anger, Sad, Disgust, Fear로 구성된 유명한 미국 드라마 'Friends'에서 추출한 영문 감성 분류 데이터셋이다. Emodya는 MELD² (Multimodal EmotionLines Dataset)에 포함된 영문 대화형 데이터 셋이다. 이는 Neutral, Joy, Anger, Surprise, Sadness, Disgust, Fear로 구성된 대화 데이터셋이다.

3.2 실험 환경

본문에서 제시한 방법론을 적용한 실험 결과를 표 1로 나타내었다.

Dataset	Emotions		Friends		Emodya	
Metric	Acc	F1	Acc	F1	Acc	F1
Cross Entropy	94.13	94.15	74.34	72.76	62.93	61.10
Vanilla SCL	94.46	94.50	75.00	72.95	62.67	58.21
Neg	94.46	94.47	75.45	73.38	63.14	59.52
Neg+Cur	94.35	94.40	75.70	73.21	63.92	59.64
Pos+Cur	94.68	94.70	75.70	73.67	64.05	60.62

표 1. 본문에서 제시된 방법 적용에 따른 성능 비교

실험 결과, Emotions 데이터 셋에서는 정확도, Weighted-F1 score가 Vanilla SCL과 비교해 각각 0.22%p, 0.2%p 상승했다. Friends에서는 각각 0.7%p, 0.72%p 상승했으며, Emodya에서는 1.33%p, 2.41%p 상승했다.

Vanilla SCL의 정답률이 94.46%로 데이터셋 중 가장 높은 Emotions에서 가장 적은 향상이 이루어졌고, Vanilla SCL의 정답률이 62.67%로 낮은 편인 Emodya에서 가장 높은 성능 향상이 이루어졌다.

Emotions는 정답률이 다른 데이터셋에 비해 높은 편이기에, 구분이 어려운 Hard-negative의 비중이 다른 데이터셋에 비해 적을 것으로 보인다. 그렇기에 우리는 본문에서 제시한 학습 방법의 효과가 Emotions에서 다른 데이터셋에 비해 약할 것이라고 추정한다. 반대로, Emodya에서는 Vanilla SCL의 정답률이 낮기에 Hard-negative의 비중이 많을 것이라 추정할 수 있으며, 그렇기에 Hard-negative를 고려한 본문에서 제시한 방법이 더욱 뛰어난 성능

¹ <https://sites.google.com/view/emotionx2019/datasets>

² https://github.com/declare-lab/MELD/tree/master/data/MELD_Dyadic

향상을 거둔 것으로 보인다.

또한, 모든 데이터 셋에서 Positive pair에 커리큘럼 러닝을 적용한 방식이 Negative pair에 커리큘럼 학습을 적용한 것보다 효과적이었다. 이는 Hard-negative로 인한 손실 함수의 값을 크게 하는 것보다, 일부 Positive pair로 인정하는 것이 작업 난이도의 설정에 더 적합하기에 커리큘럼 학습이 더 효과적이었기 때문으로 보인다.

IV. 결론 및 향후 연구 방향

본 연구에서는 감정 인식 작업에서 감정 라벨 간의 유사성을 고려한 가중치를 주는 방식으로 SCL Loss를 변형하였다. 제한한 방법을 3개 데이터셋에 대해 실험한 결과, 변형 이전과 비교해 정확도와 Weighted-F1 score에서 평균 0.75%p, 1.11%p의 성능 향상을 확인하였다. 향후 연구는 본문에서 제시된 방법을 라벨 간의 상관관계를 가지는 작업의 분류문제에 확장해 적용할 수 있을 것이다. 추후 연구에서는 본문의 방법을 더욱 발전시켜 맥락을 고려한 대화 내 감정인식 작업에 확장하고자 한다.

Acknowledgements

이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(NRF-2022R1C1C1008534)과 정보통신기획평가원의 지원(2021-0-01341, 인공지능대학원지원(중앙대학교))의 지원을 받아 수행된 연구임.

참고문헌

- [1] Yoon Kim. "Convolutional Neural Networks for Sentence Classification." In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), p. 1746-1751, 2014.
- [2] Pengfei Liu, Xipeng Qiu, and Xuanjing Huang. "Recurrent Neural Network for Text Classification with Multi-task Learning." In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI'16), pp. 2873-2879. 2016.
- [3] Jacob Devlin, Ming-Wei Chang, Kenton Lee and Kristina Toutanova. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." In North American Association for Computational Linguistics (NAACL), pp. 4171-4186, 2019.
- [4] Florian Schroff, Dmitry Kalenichenko and James Philbin. "FaceNet: A Unified Embedding for Face Recognition and Clustering." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 815-823, 2015.
- [5] Kihyuk Sohn. "Improved Deep Metric Learning with Multi-class N-pair Loss Objective." In Proceedings of Advances in Neural Information Processing Systems, pp. 1857-1865, 2016.
- [6] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu and Dilip Krishnan. "Supervised Contrastive Learning." In Proceedings of Advances in Neural Information Processing Systems, pp. 18661-18673, 2020.
- [7] James A. Russell. (1980). "A Circumplex Model of Affect." Journal of Personality and Social Psychology, 39(6), pp. 1161-1178, 1980.
- [8] Shaoling Jing, Xia Mao and Lijiang Chen. "Automatic Speech Discrete Labels to Dimensional Emotional Values Conversion Method." IET Biometrics, 8(2), pp. 168-176, 2019.
- [9] Yang Lin, Shen Yi, Mao Yue and Cai Longjun. "Hybrid Curriculum Learning for Emotion Recognition in Conversation." In Proceedings of the AAAI Conference on Artificial Intelligence, pp. 11595-11603, 2022.
- [10] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens and Zbigniew Wojna. "Rethinking the Inception Architecture for Computer Vision." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2818-2826, 2016.
- [11] Elvis Saravia, Hsien-Chi Toby Liu, Yen-Hao Huang, Junlin Wu and Yi-Shin Chen. "{CARER}: Contextualized Affect Representations for Emotion Recognition." In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 3687-3697, 2018.