

결합 베이지안 모델 기반 얼굴인식을 위한 온라인 순간 학습

곽하늬¹, 이충연¹, 이범진¹, 장병탁^{1,2}

¹서울대학교 컴퓨터공학부, ²서울대학교 협동과정 뇌과학전공

{hnikwak, swyoon, chhan, btzhang}@bi.snu.ac.kr

One-shot Online Learning for Joint Bayesian Model-based Face Recognition

Hanock Kwak¹, Chung-Yeon Lee¹, Beom-Jin Lee¹, Byoung-Tak Zhang^{1,2}

¹School of Computer Science & Engineering, Seoul National University

²Interdisciplinary Program in Neuroscience, Seoul National University

요 약

딥러닝 기술의 발전으로 얼굴을 비교하기 위한 좋은 특징 벡터를 추출할 수 있게 되었고 이에 따라 얼굴검증의 성능이 비약적으로 향상되었다. 그러나 얼굴검증 알고리즘들은 아직 1-vs-1 비교에 초점이 맞추어져 있으며 지속적으로 들어오는 무표지(unlabeled) 데이터를 온라인으로 학습하는 방법이 널리 연구되어 있지 않다. 얼굴검증을 위한 비교 데이터가 하나보다 많은 상황이 일반적이고 실제 환경에서 새로운 얼굴 영상에 지속적으로 적응해야 안정적인 성능을 유지하는데 유리하다. 본 연구는 얼굴검증에 높은 성능을 내는 결합 베이지안 모델을 확장하여 확률적인 추론에 기반한 n-vs-1 비교 모델을 설계하고 이를 이용하여 새롭게 들어오는 무표지 데이터를 온라인으로 학습하는 방법을 제안한다.

1. 서 론

얼굴인식은 주요 응용 분야에서 개인마다 학습 데이터가 적은 경우가 많고, 분류 대상이 많기 때문에 일반적인 분류 모델을 사용하는데 한계가 있다. 그래서 업계의 얼굴인식 시스템은 등록된 얼굴 영상과 주어진 얼굴 영상을 비교하여 동일한 사람인지 알아내는 얼굴검증 알고리즘이 주를 이루고 있다. 얼굴검증은 두 얼굴이 같은지 다른지에 대한 여부를 확인하는 알고리즘이다.

얼굴인식 과정은 주어진 영상에서 얼굴영역을 찾아내는 얼굴검출, 얼굴의 핵심적인 특징을 추출하는 특징추출, 그리고 얼굴검증 과정으로 진행된다. 본 연구에서는 얼굴검증 과정에 대해서 기술한다.

기존의 얼굴검증 관련 연구는 두 특징벡터의 차이, 내적, 코사인 유사성(cosine similarity) 등을 Support Vector Machine(SVM)으로 분류하는 방법[3,5,6]과 결합 베이지안 모델[1,3]을 사용하는 방법 등이 있다. 그러나 이러한 방법들은 주로 1-vs-1 비교에 초점이 맞추어져 있고 n-vs-1 비교로 일반화하기 어렵다. 본 연구에서는 결합 베이지안 모델[1]을 1-vs-1에서 n-vs-1 비교로 일반화하여 비교 성능을 높였다.

또한 본 연구에서는 실시간으로 새롭게 들어오는 무표지(unlabeled) 얼굴 영상을 활용하여 정확도를 높이는 간단한 방법을 소개한다. 초기 학습 데이터를

기반으로 무표지 얼굴 영상을 분류하고 추정되는 사람의 학습 데이터 집합에 이 얼굴 영상을 추가한다. 이 얼굴 영상이 잘못 분류되는 경우가 있지만 초기에 얼굴검증 정확도가 높기 때문에 평균적으로 성능이 증가한다.

2. 결합 베이지안 모델

결합 베이지안 모델[1]은 얼굴의 특징 벡터가 얼굴의 내재적 특성과 외재적 특성의 합으로 나타난다는 가정에서 출발한다. 내재적 특성은 모든 사람들이 공통적으로 가지는 특성이고, 외재적 특성은 사람들을 구분 지을 수 있는 특성을 의미한다. 임의의 특징 벡터 x 는 식 1과 같이 외재적 특성 μ 와 내재적 특성 ε 의 합으로 나타난다고 가정한다.

$$x = \mu + \varepsilon \quad (1)$$

μ 와 ε 이 정규 분포를 따른다고 가정한다. 가정에 따라 식 2와 같이 정리된다.

$$\begin{aligned} \mu &\sim N(0, S_\mu) \\ \varepsilon &\sim N(0, S_\varepsilon) \end{aligned} \quad (2)$$

$$x \sim N(0, S_\mu + S_\varepsilon)$$

여기서 S_μ 와 S_ε 는 얼굴의 내재적 특성과 외재적 특성의 공분산 행렬이며, 얼굴검증 이전에 다량의 얼굴

영상 데이터를 통해 학습시킨다.

최종적으로 1-vs-1 얼굴 유사도 측정을 위해 두 얼굴의 특징벡터 x_1, x_2 에 대해서 식 3와 같이 $r(x_1, x_2)$ 을 구한다.

$$r(x_1, x_2) = \log \frac{p(x_1, x_2|H_I)}{p(x_1, x_2|H_E)} \quad (3)$$

여기서 가정 H_I 는 x_1 과 x_2 가 같은 사람의 얼굴이라는 가정이고, H_E 는 반대로 다른 사람의 얼굴이라는 가정이다. 따라서 두 얼굴이 같을 확률이 더 높으면 $r(x_1, x_2)$ 이 양수가 되고 반대인 경우 음수가 된다. 두 확률 $p(x_1, x_2|H_I)$ 와 $p(x_1, x_2|H_E)$ 는 정규 분포를 따르므로 식 4과 같이 나타낼 수 있다.

$$p(x_1, x_2|H_I) \sim N(0, \Sigma_I) \quad (4)$$

$$p(x_1, x_2|H_E) \sim N(0, \Sigma_E)$$

여기서 공분산 Σ_I 와 Σ_E 는 두 얼굴에서 μ 와 ε 의 독립 관계에 따라 S_μ 와 S_ε 의 식으로 식 5과 같이 나타낸다.

$$\Sigma_I = \begin{bmatrix} S_\mu + S_\varepsilon & S_\mu \\ S_\mu & S_\mu + S_\varepsilon \end{bmatrix} \quad (5)$$

$$\Sigma_E = \begin{bmatrix} S_\mu + S_\varepsilon & 0 \\ 0 & S_\mu + S_\varepsilon \end{bmatrix}$$

최종적으로 $r(x_1, x_2)$ 는 식 6과 같이 구해진다.

$$2r(x_1, x_2) = x_1^T A x_1 + x_2^T A x_2 - 2x_1^T G x_2 + \log \frac{|\Sigma_E|}{|\Sigma_I|} \quad (6)$$

$$A = (S_\mu + S_\varepsilon)^{-1} - (F + G)$$

$$\begin{bmatrix} F + G & G \\ G & F + G \end{bmatrix} = \begin{bmatrix} S_\mu + S_\varepsilon & S_\mu \\ S_\mu & S_\mu + S_\varepsilon \end{bmatrix}^{-1}$$

3. n-vs-1 결합 베이지안 모델

본 연구에서는 1-vs-1 결합 베이지안 모델을 n-vs-1로 확장한다. n 개의 등록된 얼굴 데이터를 $X = [x_1^T, x_2^T, \dots, x_n^T]^T$, 이에 대한 비교 대상을 x 라고 할 때, 식 7와 같이 X 와 x 사이의 유사도 $r(X, x)$ 을 구할 수 있다.

$$r(X, x) = \log \frac{p(X, x|H_I)}{p(X, x|H_E)} \quad (7)$$

마찬가지로 두 확률 $p(X, x|H_I)$, $p(X, x|H_E)$ 은 식 4처럼 정규 분포를 따르고 μ 와 ε 의 독립 관계에 따라 식 8과 같이 공분산을 구할 수 있다.

$$\Sigma_I = \begin{bmatrix} S_\mu + S_\varepsilon & S_\mu & \dots & S_\mu \\ S_\mu & S_\mu + S_\varepsilon & \dots & S_\mu \\ \vdots & \vdots & \ddots & \vdots \\ S_\mu & S_\mu & \dots & S_\mu + S_\varepsilon \end{bmatrix}, \quad (8)$$

$$\Sigma_E = \begin{bmatrix} S_\mu + S_\varepsilon & S_\mu & \dots & S_\mu & 0 \\ S_\mu & S_\mu + S_\varepsilon & \dots & S_\mu & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ S_\mu & S_\mu & \dots & S_\mu + S_\varepsilon & 0 \\ 0 & 0 & \dots & 0 & S_\mu + S_\varepsilon \end{bmatrix}$$

최종적으로 $r(X, x)$ 는 식 9처럼 S_μ 와 S_ε 의 식으로 나타낸다.

$$2r(X, x) = \left(\sum_i x_i^T \right) (G' - G) \left(\sum_i x_i \right) + x^T A x - 2x^T G \left(\sum_i x_i \right) + \log \frac{|\Sigma_E|}{|\Sigma_I|} \quad (9)$$

$$G = -((n+1)S_\mu + S_\varepsilon)^{-1} S_\mu S_\varepsilon^{-1}$$

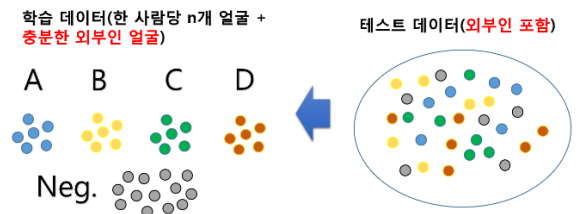
$$G' = -(nS_\mu + S_\varepsilon)^{-1} S_\mu S_\varepsilon^{-1}$$

$$A = (S_\mu + S_\varepsilon)^{-1} - (F + G)$$

식 9에서 S_μ 와 S_ε 은 사전 학습이 완료된 후에는 변하지 않는 값이고, 이 둘로부터 계산되는 A , G , G' , $\log \frac{|\Sigma_E|}{|\Sigma_I|}$ 도 재사용이 가능하다. 또한 $(\sum_i x_i^T)$ 도 미리 계산해 둘 수 있고, 새로운 얼굴 영상이 등록되는 경우 추가적으로 합산만 하면 된다. 따라서 $r(X, x)$ 의 계산은 시간복잡도 $O(n^2)$ 으로 간단하고 효율적이다.

4. 실험 내용 및 결과

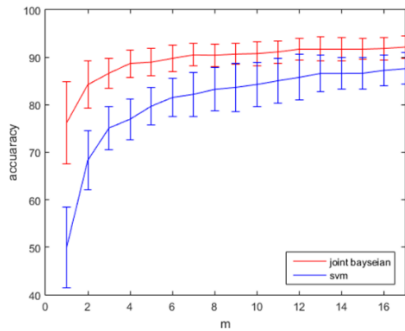
4.1 열린 환경에서의 순간 학습



[그림 4.1-1] 열린 환경에서의 순간 학습

네 명으로 구성된 가정 환경을 가정하자. 주로 이 가족 구성원들이 얼굴인식을 사용하지만 가끔 외부인도 있는 열린 환경을 가정한다. LBP와 Gabor feature를 차원 축소법으로 압축한 600차원 특징 벡터를 사용하였고, 데이터는 Labeled Faces in the Wild(LFW)를 사용하였다. 여기서의 학습데이터는 한 사람당 n 개로 고정하고, n 에 따른 인식 정확도를

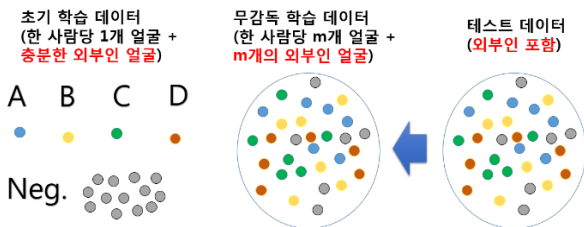
측정하였다. 정확도를 측정하는데 사용된 테스트 데이터에는 사람마다 균등한 수의 얼굴이 포함되었다. 비교 모델로 대표적인 분류 모델인 SVM을 사용하였다. [그림 4.1-2]는 n 에 따라 결합 베이지안 모델과 SVM 모델의 성능을 비교하고 있다. 꺾은 선 사이에 있는 막대는 100회 반복 실험에서 측정된 표준편차를 나타낸다.



[그림 4.1-2] 닫힌 환경에서의 순간 학습 실험 결과

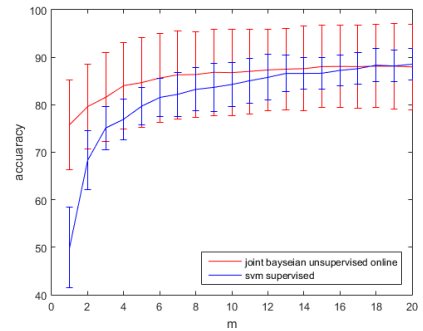
[그림 4.1-2]의 실험 결과에서 결합 베이지안 모델이 SVM에 비해 더 좋은 성능을 내고 있다. 또한 n 이 늘어남에 따라 정확도가 90% 정도까지 향상되고 표준편차가 줄어들어 안정적인 성능에 수렴한다. 여기서 n -vs-1 비교가 1-vs-1 비교보다 더 높은 정확도를 나타내고 있다.

4.2 열린 환경에서의 무감독 온라인 학습



[그림 4.2-1] 열린 환경에서의 무감독 온라인 학습

본 실험에서는 초기에 사람당 1개의 학습 데이터로 시작해서 순차적으로 들어오는 무감독 학습 데이터를 받아들인 다음 테스트 데이터를 분류한다. 무감독 학습 데이터는 얼굴의 소유주 정보가 없는 학습 데이터이다. 주어진 무감독 학습 데이터를 가장 높은 확률로 추정되는 사람의 학습 데이터 집합에 추가하는 방식을 사용하였다. 초기 학습 데이터의 편향에 따라 최종적인 성능에 악영향을 줄 수 있지만 실험적으로 이전 실험의 SVM이 낸 성능보다 좋은 결과가 나왔고 m 이 증가함에 따라 성능이 향상되는 현상을 발견하였다. [그림 4.2-2]은 이 실험의 결과를 나타낸다.



[그림 4.2-2] C 시나리오 실험 결과

5. 결론 및 논의

결합 베이지안 모델 기반의 n -vs-1 비교 방법은 통계학적으로 잘 정리된 비교법으로서 1-vs-1에 비해 더욱 높은 정확도를 보인다. 또한 새로 인식된 얼굴을 등록 데이터에 계속 추가하면서 지속적으로 성능을 향상시킨다. 이 비교 방법은 n 이 커짐에 따라 정확도가 향상되며 잘못된 얼굴이 섞여있어도 다른 올바른 얼굴들에 의해 보완된다. 일반적인 얼굴인식 응용에서 사전에 많은 얼굴 영상을 등록하기 어렵기 때문에 실제 작동 상황에서 지속적으로 등록 데이터를 늘려가는 것이 중요하다.

감사의 글

이 논문은 2015년도 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원(R0126-15-1072-SW스타랩, 10044009-HRI.MESSI)을 받아 수행된 연구이며, 삼성전자의 지원을 일부 받았음.

참고문헌

- [1] Chen, Dong, et al. "Bayesian face revisited: A joint formulation." Computer Vision-ECCV 2012. Springer Berlin Heidelberg, 2012. 566-579.
- [2] Sun, Yi, Xiaogang Wang, and Xiaoou Tang. "Deeply learned face representations are sparse, selective, and robust." In Computer Vision and Pattern Recognition (CVPR), 2015
- [3] Taigman, Yaniv, et al. "Deepface: Closing the gap to human-level performance in face verification." Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on. IEEE, 2014.
- [4] Sun, Yi, et al. "Deep learning face representation by joint identification-verification." Advances in Neural Information Processing Systems. 2014.
- [5] Hu, Junlin, Jiwen Lu, and Yap-Peng Tan. "Discriminative deep metric learning for face verification in the wild." Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on. IEEE, 2014
- [6] Chen, Dong, et al. "Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification." Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on. IEEE, 2013.
- [7] B.T. Zhang. "An incremental learning algorithm that optimizes network size and sample size in one trial." Neural Networks, 1994. IEEE World Congress on Computational Intelligence, 1994 IEEE International Conference on. Vol. 1. IEEE, 1994.