Networks with emotions

Mihaela Rosca
Imperial College London, Department of Computing

Abstract

Facial emotion recognition from images has been a field of intense research for decades, its difficulty coming from its interdisciplinary nature. The problem can be divided in two main tasks: feature extraction and emotion classification. We tackle both using deep belief networks: unsupervised pre-training for feature learning and supervised training for classification. In addition, we define a probabilistic model that can learn if two human subjects display the same emotion without requiring emotion classification. We complement our theoretical work with a fast open source GPU implementation of all models used. The results obtained on our experiments are positive: a classification accuracy of 99.3% on the Multi-PIE dataset makes our results comparable with the state of the art.

Motivation

The multitude of practical applications which arise from emotion recognition has driven the research in the field. A system able to accurately detect emotions can be used in judicial settings (recognizing if criminal subjects are genuine or not), in entertainment (interactive games that can increase pace if the player is bored and decrease it if the player is overwhelmed), in medicine (teaching autistic children to detect emotions) and ultimately in AI (any robot that is supposed to live with humans should be able to identify our emotional status).

Deep belief networks

Deep belief networks (DBNs) are probabilistic generative models that efficiently learn the probability distributions between the observed data (the visible layer) and latent variables (the hidden layers) through greedy layer-wise pre-training.
Training a deep belief network has two phases:

- Unsupervised pre-training for learning features
- Supervised learning to learn the labels (by adding a softmax layer and using backpropagation)

Unsupervised pre-training

DBNs learn features from the data in a greedy layer wise fashion, by training a Restricted Boltzmann machine (RBM) for each layer. RBMs are probabilistical models that aim to learn the data distribution during training. Parameters are learned with Contrastive Divergence, a computationally efficient algorithm that aims to increase the log likelihood of the data under the model distribution.

Figure: A Restricted Boltzmann machine.

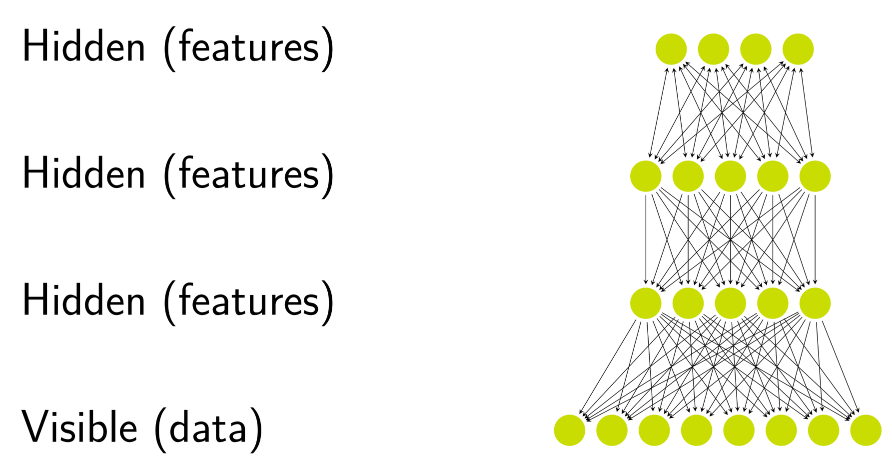

Figure: A deep belief network.

Neural networks methods

We used the following techniques to increase classification accuracy and efficiency of training:

- Dropout (2012)
- Rmsprop (2013)
- Rectified linear units (2010)
- Our new extension for sparsity constraints (2014)

The data

Figure: The 7 emotions from the Cohn-Kanade dataset

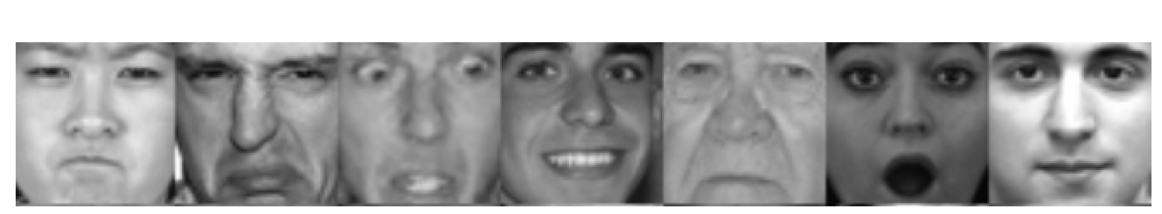

Figure: The 7 emotions from the Kaggle dataset

Experiments

We assessed the performance of our model by testing it on multiple datasets. The dataset from a Kaggle competition is the hardest to classify, due to variation in how emotions are depicted and the lack of face alignment.

Dataset	Nr. emotions	Size	Accuracy
Cohn-Kanade	6	400	81.0 %
Multi-PIE	6	28000	99.3 %
Kaggle	7	3000	70%

Table: Random splits results

The Multi-PIE dataset provides more meta-data for each image: subject, illumination and pose. We used this information to devise experiments that test the robustness of our network: we trained it with instances which have certain properties, and tested it ones that do not. For example, we trained the network with faces in four out of the five poses that we used and tested with the fifth. This allows us to assess the generalization capability of our model. We tested for noise robustness by depriving the network of pixel information in squares of various sizes.

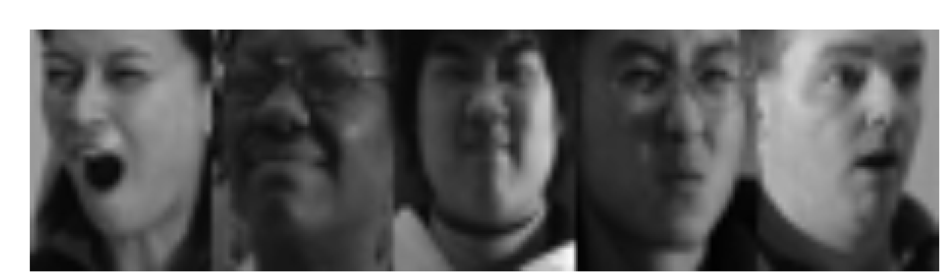

Figure: The 5 poses from the Multi-PIE dataset

Experiment	Accuracy
Random data splits	99.3 %
Different subjects	91.7 %
Different illuminations	94.8 %
Different poses	36.5 %
Missing data 5×5	91.5%
Missing data 10×10	65 %

Table: Experiments performed on the Multi-PIE database

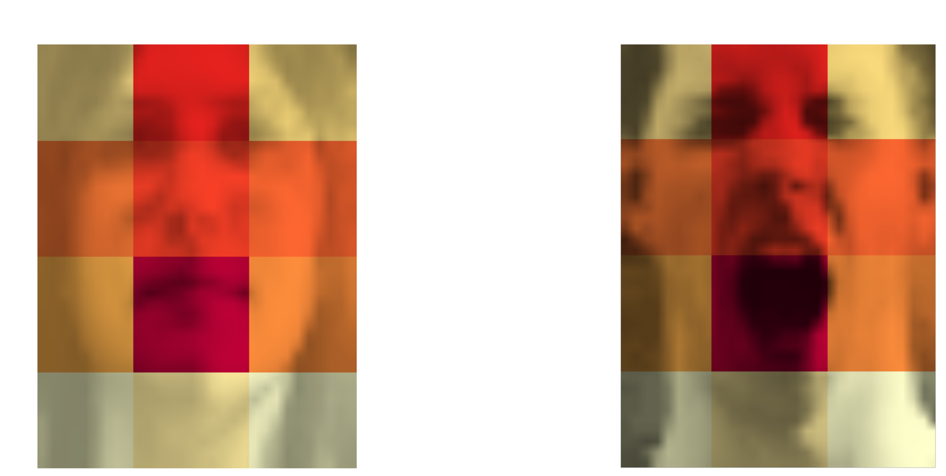

Figure: Facial zones according to emotion classification error

Emotion similarity

A new model was defined in order to assess the capability of neural networks to distinguish between emotions. The model is fed two input images depicting faces and outputs the probability that the two subjects are displaying the same emotion. For this experiment we obtained an accuracy of 90.5 % if the subjects depicted in the two images are the different, and 92.5% if we distinguish between emotions depicted by the same person.

Figure: Example of inputs pairs, aligned vertically

Conclusion

We have shown that unsupervised pre-training in deep belief networks provides an excellent framework for feature extraction from images of faces. When measuring classification accuracy, we obtained state of the art results on three datasets. Using a novel experiment we have shown that we can distinguish between emotions without requiring individual emotion classification.

References

[1] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006.

Extra information

- Report: http://www.doc.ic.ac.uk/teaching/distinguished-projects/2014/mrosca.pdf
- Code: https://github.com/mihaelacr/pydeeplearn
- Demo Video: http://elarosca.net/video.ogv
- Contact: mcr10@imperial.ac.uk