

論文の要旨

題目 Feature extraction via deep learning for emotions estimation
(感性推定のための Deep Learning による特徴抽出)

氏名 Motaz Sabri

Optimizing the extraction of feature sets for classification tasks is still a fundamental and challenging problem in the area of machine learning. For Mammals to perform classification tasks we use hierarchical features and efficient coding in our visual pathway. Information in visual system is encoded using distributed coding schemes and later the sparse coding is utilized. We propose multiple architectures to extract features for emotional analysis that encode the information according to a specific activation profile. We show how our models much like the visual system, can learn distributed coding in lower layers and sparse coding in higher layers for emotional analysis. Feature extraction for emotion related tasks must ensure high accuracy in locality and activation. Our models can dynamically extract features and perform classification and ranking for many commonly used datasets. As for emotional classification through facial expressions, we introduce regularization through noisy training that helped internal representation of learned features to emerge and increased sparsity of hidden units. This regularization also improved the network generalization through automatic structuration. For our ranking models that estimate emotion intensity, they extracted the sequential relationship in the temporal domain that appears due to the natural change of facial expression during an emotional transition. The output ranking score can be directly used for intensity estimation for all emotional states. This allows an early detection of an emotional transition and estimation of intensity of an emotion considering temporal information.

Beside facial expressions for emotion analysis, we proposed an eye tracking that can be used to understand person emotions through pupil movement. Through Gabor filter, we detected the intensity change of the gaze edge by applying Gabor filter and its second moment matrix. Eigenvalues inequality of this matrix with a given local intensity-based threshold defines the prevailing directions of the neighboring gradients and the amount by which these directions are coherent at a certain point. This allows accurate estimation of the gaze direction.

We find that all CNN features can be used for knowledge representation purposes both by their activation and locality, offering valuable information from single CNN feature. We also study the levels of importance for these features, and propose approaches and structures to discard most features that are invaluable for the decision making. All these insights have a direct application to the generation of CNN embedding spaces for the field of emotion analysis.