

Abstract

The concept of entropy and related measures has been applied in learning systems since the 1980s. Several researchers have applied entropic concepts to independent component analysis and blind source separation. Several previous works that use entropy and mutual information in neural networks are basically related to prediction and regression problems.

In this thesis we use entropy in two different perspectives: first as a cost function in neural networks (multi-layer perceptrons - MLP) for supervised classification problems; and second, as the basis for a new entropic measure for unsupervised classification (clustering).

Related with the new entropic cost function we show in the present work how to use Rényi's quadratic entropy of the errors between the output of an MLP and the desired targets as a cost function in classification problems: the Error Entropy Minimization Algorithm (EEM). We present several optimization procedures for this algorithm, namely: how to use an appropriate adaptive learning rate; how to tune the smoothing parameter when performing entropy gradient computation; and, the conditions to apply a batch-sequential algorithm for training with the EEM algorithm.

Regarding unsupervised classification, we present a clustering algorithm based on a new entropic dissimilarity matrix. This matrix is the basis for a new clustering process, based on layered entropic subgraphs, which we call LEGClust.

We also used this algorithm to perform task decomposition in modular neural networks for classification problems.

Resumo

Desde os anos 80 que o conceito de entropia e medidas relacionadas se aplicam a sistemas de aprendizagem tendo-se assistido à aplicação destes a problemas de *independent component analysis* e *blind source separation*. Trabalhos posteriores que

usam a entropia e a informação mútua em redes neuronais estão fundamentalmente relacionados com problemas de regressão e predição.

No trabalho que aqui apresentamos, usamos a entropia de duas formas distintas: primeiro, como função de custo em redes neuronais (perceptrão multicamada) para problemas de classificação; segundo, como base para uma nova medida entrópica para classificação não supervisionada (*clustering*).

Relativamente à nova função de custo, mostramos como usar a entropia quadrática de Rényi dos erros entre a saída do perceptrão multi-camada e as respectivas etiquetas, como função de custo em problemas de classificação: o algoritmo de minimização da entropia do erro (EEM). Apresentamos também um conjunto de optimizações para este algoritmo, nomeadamente: a adopção de uma taxa de aprendizagem variável; a afinação do parâmetro de suavização para estimação do gradiente da entropia; as condições para o uso de um algoritmo "batch-sequential" para treino das redes neuronais com o algoritmo EEM.

No que respeita à classificação não supervisionada, apresentamos um algoritmo de *clustering* baseado numa nova matriz entrópica de dissimilaridade que serve como base para um novo processo de *clustering*, ao qual chamamos LEGCLust.

Também usamos este novo algoritmo para efectuar a decomposição de tarefas para redes neuronais modulares em problemas de classificação.