

クラウドソーシングによる訓練データセット構築における最適な冗長度の検証

清水 綾女

教師あり機械学習ではクラウドソーシングを用いた訓練データの作成が広く行われている。しかし、クラウドワーカーの回答は品質が保証されないため、データの集約には工夫が必要とされる。一般的な対策として複数人から同じタスクに対する回答を得る冗長化が行われているが、クラウドソーシングにかけられる予算に上限があると考え、冗長化は回答の精度を向上させる一方で訓練データの量を減少させる。本研究では、予算、ワーカーの回答精度、学習器にとってのタスクの難易度の3変数から、クラウドソーシングの効用を最大化させる冗長度の検証をシミュレーションデータによる実験で行った。

(指導教員 若林啓)