

平成30年度 情報工学コース卒業研究報告要旨

工藤 研究室	氏 名	波 多 野 優 斗
卒業研究題目	特徴空間上におけるN近傍の属性に基づくラベル推定を利用した半教師あり能動学習	

近年、情報通信端末の普及により、端末内のデータを整理することの必要性が高まっている。これまでも、少数のラベルありデータと多数のラベルなしデータを使って学習を行う半教師あり学習と、質の高いデータにラベルを付与できるようなデータを選択する能動学習を組み合わせた研究は行われてきた。しかし、識別率がRandom Samplingに劣る、汎用性が低いなど、問題が多い。本研究では、半教師あり学習のひとつであるSemi-Supervised Support Vector Machine (S3VM) を対象とした能動学習手法を提案する。

提案手法は、S3VMによる分類器の学習、能動学習、ラベル推定の3段階を1ステップとしてラベル付与を行う。提案手法の処理を図1に示す。能動学習では決定境界からの距離を基にラベル付与データを選択し、ユーザーがラベルを付与する。ラベル推定では、ラベルなしデータのうち、特徴空間上で距離が近い規定個のラベルありデータについてのラベルがすべて同一であるものに対して、一時的にシステムがラベルを付与する。これらのラベルは再びラベル推定を実行する時は付与し直される可能性がある。ラベル推定には、ラベルありデータの不足を補うことで分類器を正確なものにし、能動学習の精度を向上させる効果がある。

提案手法とラベル推定の有効性の検証のため、評価実験を行った。今回の実験では、旅行時に撮影した写真を整理することを想定し、入力を建築物と料理の写真に限定した。実験は、Random Sampling, 提案手法(ラベル推定なし), 提案手法(ラベル推定あり)の3つの手法により、5分割交差検証で評価した。ラベルありデータの割合を4%から開始し、既定の割合に達する毎に、テストデータによる評価を行った。結果を図2に示す。提案手法はラベル推定のあり・なしに関わらず学習の初期段階では良い結果を示したが、その後はRandom Samplingと同程度の結果を示した。

今後の課題としては、ラベルありデータが10%を越えた時の識別率を改善するためにラベル付与データの選択基準の改善があげられる。また関連研究の結果に見られるように、撮影時間などの潜在情報を利用することで学習の性能の向上が期待できると考える。

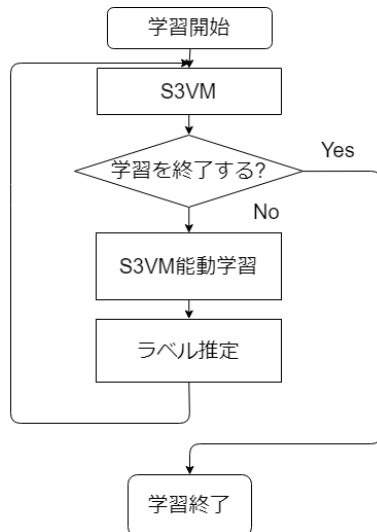


図1：提案手法の処理

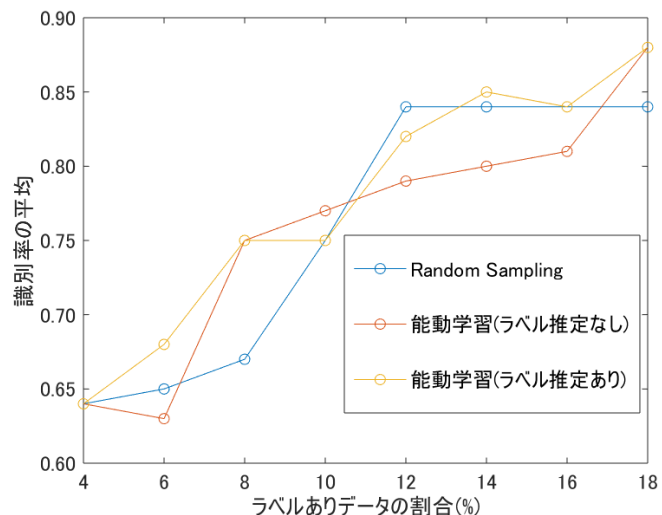


図2：各割合での識別率の平均