

平成15年度 情報工学コース卒業研究報告要旨

大西 研究室	氏 名	不 破 洋 平
卒業研究題目	複数代表点による部分教師付き学習 - 類似画像検索への応用 -	

背景と目的

近年、一般家庭へのコンピュータ、デジタルカメラ等の普及により、大量の画像データを収集、蓄積することが一般的になりつつある。そのため、一般のユーザでも容易な操作によって、画像を分類、検索できるシステムが求められている。このようなシステムにおいては、可能な限りユーザの介在なしに画像の分類・検索が可能であることが求められる。しかし、多様なカテゴリの画像が混在する画像データベースにおいて、完全に自動分類を行うことは困難であり、従来は画像カテゴリの人手によるラベル付けが行われてきた。

そこで本研究では、少量の画像のカテゴリを教師として指定するだけで、高精度なカテゴリ分類を実現することを目的として、部分教師付き学習を、類似画像検索に応用することを検討した。

学習の手法

本研究で利用する、部分教師付き学習とは、カテゴリが未知なデータとカテゴリが既知であるデータが混在した状態で用いることが可能な学習法であり、教師なし学習に関する評価量と教師あり学習に関する評価量の同時最適化を行うものである。これにより、カテゴリ未知のデータとカテゴリ既知のデータの双方を利用して、より良い識別結果が得られることが期待される。また1カテゴリ多代表点の最近傍識別器を構成することで、複雑なデータ分布にも対応できると期待される。

この学習を類似画像検索に用いる方法として、まず教師なし学習を実行し、その結果をユーザが各代表点ごとに、自分の意図する画像が分類できているか判断する。その結果に対して、幾つかの画像データのカテゴリを指定し、部分教師付き学習を行う。これにより検索精度の向上を目指した。

実験と結果

実画像データを用いて、教師なし学習、教師あり学習、部分教師付き学習の比較実験を行った。画像データの特徴量は、色相のヒストグラムとウェーブレット変換係数に、主成分分析により次元を削減したデータを学習に用いた。実験方法は、まず教師なし学習を行いその結果に基づいて少数のデータのカテゴリを指定し、その後部分教師付き学習、または教師あり学習を行い、識別率の変化を調べた。

実験の結果、カテゴリ間でデータに偏りがない場合には、部分教師付き学習による識別率の改善は確認できなかった。カテゴリ間でデータ数に偏りがある場合には、教師なし学習では少数データのカテゴリにおいて識別率の低下が見られたが、部分教師付き学習では識別率が改善された。

また、教師データ数の変化に対しては、教師データ数が少ない場合は部分教師付き学習は教師なし学習より識別率が低い場合が見られたが、教師データ数の増加に伴って、教師なし学習よりも高い識別率を示すことが確認された。

カテゴリのデータ数に比例した教師データ数を与えた場合(部分教師(教師数比例))、各カテゴリ同程度の教師データ数を与えた場合(部分教師(教師数同数))の部分教師付き学習と、教師なし学習について、教師データ総数を変化させた時の識別率を図1に示す。

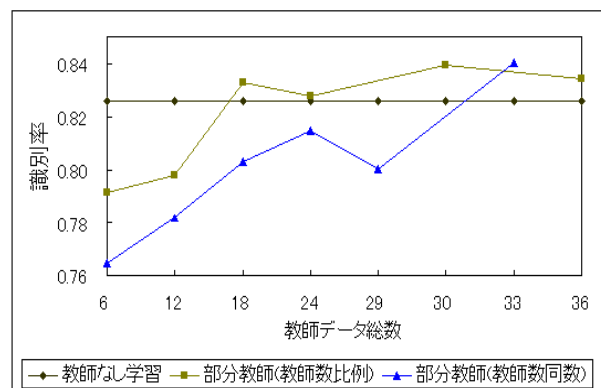


図1 教師データ数と識別率の関係
(カテゴリ数3, データ総数120)