

## 平成 19 年度 情報工学コース卒業研究報告要旨

大西 研究室	氏 名	太 田 創		
卒業研究題目	GMM による部分教師付き学習 ～ 類似画像検索への応用 ～			
<p><b>背景と目的</b>            近年、デジタルカメラ等の普及により、コンピュータ利用者が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。このため、一般の利用者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。従来は人手によって画像カテゴリが入力されてきたが、利用者によって趣味・嗜好は異なり、このようなアプローチは画像数やカテゴリ数が増大するにつれて、現実的とは言えなくなる。            そこで本研究では、教師となる少数の画像のカテゴリを指定するだけで、高精度なカテゴリ分類を実現することを目的とし、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。</p> <p><b>学習の手法</b>            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、その双方のデータを考慮した評価量を最適化する学習法である。本研究では、学習データの分布を GMM(混合ガウス分布モデル) で表現した。そして、カテゴリが既知であるデータを考慮した EM アルゴリズムにより GMM に関するパラメータを学習した。また 1 カテゴリを複数個のガウス分布でモデル化することで、複雑なデータ分布にも対応できると期待される。</p> <p><b>実験と結果</b>            人工データと実画像データを用いて、教師なし学習と部分教師付き学習の識別率の比較を行った。画像データの特徴量には、色相のヒストグラムとウェーブレット変換係数を使用し、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。            実験の結果、部分教師付き学習のほうが教師なし学習より識別率が良い場合が多く見られ、僅かな数でも教師を与えることによって、学習結果を改善することが可能であることが示された。ガウス分布数を変えた実験における両学習の誤識別率を表 1、表 2 に示す。            また、今回利用した部分教師付き学習アルゴリズムと比較するために、EM アルゴリズムの初期値として、ガウス分布の平均ベクトルに教師データを与える手法(学習アルゴリズムは教師なし学習と同一)とも比較実験を行った。その結果、全体的には、両手法とも同程度の誤識別率となった。</p>				
表 1 . ガウス分布数を変える実験の部分教師付き学習の誤識別率				
各カテゴリのガウス分布数	平均値	標準偏差	最大値	最小値
3 個	0.435	0.069	0.613	0.435
9 個	0.473	0.049	0.571	0.379
15 個	0.453	0.069	0.608	0.342
表 2 . ガウス分布数を変える実験の教師なし学習の誤識別率				
各カテゴリのガウス分布数	平均値	標準偏差	最大値	最小値
3 個	0.501	0.049	0.583	0.383
9 個	0.557	0.035	0.617	0.483
15 個	0.585	0.035	0.654	0.513