



## (論文審査結果の要旨)

統計的学習において、訓練データの不足はしばしば起る深刻な問題である。これまでも、量や質の面で十分ではないデータからの学習を可能とするために、様々な手法が、提案・研究されてきた。しかし、それらの多くはモデルの推定を安定させる効果はあるものの、得られたデータの偏りのために興味のある現象全体を説明できない、という問題に対しては有効ではない。

本論文では、データの収集が比較的容易な他の情報源を利用することで、学習における訓練データ不足の解決を試みている。情報不足が深刻となる典型的な二種類の情報処理課題を対象とし、そこで利用可能な情報源、及びその効果的な利用法について考察している。本論文の主な成果は以下のように要約される。

1. 分類問題において、人手でクラスラベルを付与するコストが多いため、訓練データが不足し、学習が適切に行われないという問題が存在する。データ収集のコストの大部分は人手によるラベル付与によるものであり、ラベル付けがなされていない「ラベル無しデータ」が豊富に存在するが、それらはこれまで系列データのモデルでは利用されていなかった。そこで、本論文では広く使われているモデルである隠れマルコフモデルにおいてラベル無し系列データの利用法を提案し、ジェスチャー及び音声の識別実験によりその有効性を示している。
2. 情報検索の中でも、テキストによる問い合わせで画像を検索するなど、問い合わせと対象データのメディアが異なるクロスメディア検索は、検索対象に付与されたアノテーション（単語による説明）が少ないために困難な問題となる。特に、単語間関係を構築・利用することが情報検索においては必須となるが、少数の単語からでは、それらをモデルとして適切に推定することは容易ではない。そこで、本論文ではアノテーションと対になった画像の情報を利用することで、情報不足を補完する方法を提案している。写真画像を対象とした情報検索実験により、提案手法の有効性を示している。

本論文の成果は、統計的学習における基本的課題である情報不足の問題に対し、複数情報源を利用するという対処法を、系列データの分類問題及びクロスメディア情報検索問題において具体的なモデル及びアルゴリズムとして示したことである。これらは、知的情報処理モデルの統計的学習、特に高精度のパターン認識及び情報検索の実現において、学術上、および、応用上の寄与が十分である。よって、博士（理学）の学位論文として価値のあるものと認める。