

【招待講演】

分布のカーネル埋め込みに基づく Bag-of-Words データのための潜在変数モデル

吉川 友也

千葉工業大学

特徴の多重集合で表現されるデータは Bag-of-Words(BoW)データと呼ばれ、NLP や画像処理等のデータ表現として用いられる。我々は BoW データに対して、(1) 各特徴を潜在ベクトルで表現し、(2) 各 BoW データを潜在ベクトルの分布とみなし、(3) 潜在ベクトルの分布を「分布のカーネル埋め込み」によってノンパラメトリックに表現する、新しいデータ表現法を考案した。本講演では、このアイディアの詳細を説明するとともに、このデータ表現法に基づく分類・回帰・異種データ間マッチングの問題を扱う潜在変数モデルを紹介する。