

平成28年度 情報工学コース卒業研究報告要旨

金森 研究室	氏 名	飯室崇晃
卒業研究題目	非負値行列因子分解におけるロバスト性の検証	
<p>私たちの身の回りにあるデータのほとんどは、基本的に要素が非負の行列型データとして表現でき、そのデータを解析することで、データが持っている特徴を抽出することができる。例えば、顔画像データを目や口、鼻などの部位にあたる画像データに分解することが出来れば、認証システムなどに役立てることが出来る。非負値行列因子分解 (Non-negative Matrix Factorization, NMF) は、このようなデータを解析する手法として提案された。他の応用例としては、複数の音源の音声信号が混ざったデータを解析し、別々の音声信号として分離することで、それぞれの音源に対してノイズ除去をかけることができる。また、文章データ (文章における単語の出現頻度を要素とし、文書と単語の行列にしたもの) を解析して、文書クラスタリングに利用することで、文章検索に役立てることができる。しかし、実際に得られるデータにはある程度のノイズが含まれており、得られたデータが正確ではないことがほとんどである。したがってNMFアルゴリズムによって得られる解析結果は、信頼性にかけてしまう。これを解決するために、ノイズに対してロバストであるNMFアルゴリズムを解析手法として用いれば、通常のNMFアルゴリズムで得られる解析結果よりも信頼性の高いものとなる。</p> <p>NMFとは、解析対象である非負値行列 $V(n \times m)$ を、2つの低次元の非負値行列 $W(n \times k)$ と $H(k \times m)$ に分解する解析手法である。NMFアルゴリズムの目的は行列積 WH を行列 V に近似することであり、どれだけ近似できているかをコスト関数を用いて判断する。そのコスト関数を行列積 WH の関数としてみなし、W と H を変化させ、コスト関数を最小化することによって、V を分解した W と H が得られる。</p> <p>本研究では、コスト関数としてよく用いられるユークリッド距離・カルバックダイバージェンス・板倉斎藤距離、KL 距離よりも外れ値に対してロバストであるピアソン距離を用いてNMFアルゴリズムを実行した時に、どの目的関数を利用すればノイズに対してロバストであるかを検証することを目的としている。行列 W_0 と H_0 を用意し、W_0 と H_0 の行列積 WH_0 にノイズを加えたものを V としてNMFアルゴリズムを実行した時に、更新して得られた W と H が、WH_0 にどれだけ近似できているかを比較した。NMFアルゴリズムを実行して、得られた W と H の積を WH_0 と比較した結果、ノイズが大きい場合やノイズの割合が高い場合にピアソン距離をコスト関数としたNMFアルゴリズムがもっとも行列間の隔たりが少なく、ノイズに対してロバストであるという結果になった。</p>		