

## 論文内容の要旨

博士論文題目 効率的な近傍検索のためのピボット学習法

氏名 木村 学

(論文内容の要旨)

近傍検索は、与えられたクエリオブジェクトに類似したオブジェクトをデータベース中から見つけるタスクである。画像、音声、ビデオクリップ、3次元オブジェクトなどのマルチメディアデータの急速な増加とともに、その重要性は高まっている。

マルチメディアデータの類似度計算は高コストであり、多くの時間が必要なため、問い合わせ時にクエリと全データオブジェクト間の類似度計算を行うのは現実的ではない。

ピボットに基づくインデキシングは、ピボットオブジェクト集合とデータベース内の全データオブジェクト間類似度を事前に計算しておき、これを利用して問い合わせ時の類似度計算回数を減らす方法である。既存のピボットインデキシング法は、組み合わせ最適化技術を使い、データベース内のデータオブジェクトからピボットとなるオブジェクト集合の候補を選ぶ。

本博士論文では、データベース内には必ずしも存在しないオブジェクトから、ピボットオブジェクトの集合を学習する機械学習アプローチを提案し、その有効性を示す。このアプローチに基づいて二つのアルゴリズムを提案する。一方はユークリッド空間のオブジェクトに適用できるアルゴリズムである。もう一方はカーネル関数により暗に定義される特徴空間のオブジェクトに適用できるアルゴリズムである。後者のアルゴリズムは、前に述べたマルチメディアデータを含む、多様な非ベクトルデータ（構造データ）に対しても効率的な近傍検索を可能にする。

ベクトルデータや構造データ（自然言語文の構文解析木）のデータベースを用いた実験を行い、データベース内のデータオブジェクトからピボットを選ぶ、既存のインデキシング手法に比べて、提案するアルゴリズムが問い合わせ時の類似度計算回数を大幅に削減することを確認し、その有効性を示す。

氏名	木村 学
----	------

(論文審査結果の要旨)

平成22年1月22日に開催した公聴会の結果を参考に平成22年2月15日に本博士論文の審査を行った。以下のとおり、本博士論文は、提案者が独立した研究者として、研究活動を続けていくための十分な素養を備えていることを示すものと認める。

木村学は、本博士論文において、大規模データに対する近傍検索を効率化するための手法であるピボット法に関して、性能の高いピボット集合を求める手法を提案した。さらに、同手法を拡張し、カーネル関数によってデータ間の類似度が定義された空間においても適用可能になる方法を示した。また、実データおよび人工データに対して、提案した手法の有効性を実験により確認した。

本論文の貢献は、次のようにまとめることができる。

1. 従来のピボット法では、最適なピボット集合の選択は、検索対象のデータ集合の中から行われるのが普通であった。これに対して、本論文では、対象データの空間中の任意のデータ点を候補としてピボット集合を選択できる方法を示し、実験的に性能向上が達成できることを示した。
2. 近年、遺伝子データ、化合物、言語データ等、様々な複雑な構造を持つ大規模データが取り扱われるようになり、構造を考慮したデータ間の類似度の定義として、系列カーネルや木構造カーネルなど様々なカーネル関数を用いられるようになってきている。本論文では、上記の手法をカーネル関数による類似度定義にも拡張できることを示し、近傍検索の効率化を達成することを実験により確認した。

近傍検索のためのピボット法を、ピボット集合の選択法、および、カーネル関数に基づく類似度定義という2つの方向から拡張した本研究は、独創性が高く、しかも実用的であり、自然言語処理の分野において高い貢献があると評価する。

よって、本論文は、博士(工学)の学位論文として価値あるものと認める。