

## 論文内容の要旨

博士論文題目 ハブ削減による近傍法の精度改善

氏名 重藤優太郎

(論文内容の要旨)

近傍検索・k近傍分類などの近傍法における問題の一つとして特徴空間上のハブの出現 (hubness phenomenon) が注目されている。ここで言う「ハブ」とは、データセット中の多数の事例から見て最近傍 (あるいは最k近傍) となる少数の事例のことを指す。そのようなハブが出現するデータセットに対して近傍検索を行うと、様々なクエリ事例に対して同一のデータ事例 (ハブ事例) が検索結果として何度も出力される。またk近傍分類では、どのようなテスト事例を入力してもハブ事例と同じラベルが予測結果となる。このように、ハブ事例の存在は近傍法の性能低下を引き起こす一つの要因になっている。

本論文では、このハブの出現を抑制することで近傍法の性能改善を行った。

本研究では、はじめに特徴空間上でハブがなぜ出現するかを分析した。具体的には、特徴空間の次元数、もしくは、事例間距離の分散が大きくなるほどハブが出現しやすいことを示した。これまでもハブが出現する理由について議論が行われていたが、それらは特徴空間の次元数についての議論のみであり、次元数以外のデータの性質 (本研究の事例間距離の分散) とハブの関係については議論されていなかった。

次に、この分析結果に基づいてハブ削減法を提案した。具体的には、事例間距離の分散を小さくすることでハブの抑制を行った。これは単純なリッジ回帰で実現される (ただし通常のリッジ回帰とは写像方向が異なる) ので、学習が簡単であり、また負例を必要としない。

実験の結果、提案した手法は実際にハブを抑制し、それに伴い近傍法の精度が改善された。

氏名	重藤優太郎
----	-------

(論文審査結果の要旨)

平成29年1月23日に開催した公聴会の結果を参考に平成29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。

重藤優太郎は、本博士論文において、データ間の類似度に基づく特徴空間に生じるハブの問題を取り上げ、その性質の解明やハブの影響を軽減するための手法を提案した。さらに、いくつかの問題について、提案した手法に基づいて近傍法の性能向上が達成できることを示した。

本論文の貢献は、以下のようにまとめることができる。

1. 特徴空間上でハブがなぜ出現するかを分析を行い、特徴空間の次元数、もしくは、事例間距離の分散が大きくなるほどハブが出現しやすいことを示した。
2. 二言語の大規模なテキストデータと小規模な対訳語に基づいて言語間の単語の類似度を定義し、対訳語を集中する手法が提案されていたが、そこでもハブの問題による性能低下が生じていることを指摘し、ハブ軽減の手法を適用することにより、簡単なベースライン手法においても、大きな性能向上が可能であることを示した。
3. ゼロショット学習におけるハブの影響について考察し、距離空間の写像を通常とは逆に地域から定義域への写像を定義することでハブの影響が軽減されることを見出した。さらに、データが多変量正規分布に基づくとの仮定の上で、提案した写像の方向がなぜハブの軽減につながるのかを説明した。
4. k近傍法においても、上記手法がハブの軽減につながり、性能向上を実現することを示した。

このように、近傍法に基づく大規模なデータ空間によるハブの性質を詳細に解析し、種々の問題において性能向上を達成する手法を導いた本研究は、独創性が高く、実用的であり、機械学習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。