

平成 30 年 8 月 6 日

## 学 位 論 文 の 審 査 要 旨

学位論文申請者氏名：Rachelle Alvarez Rivero

論 文 題 目： “Kernel Matrix Completion” (カーネル行列補完)

### 論文の概要及び判定理由

Rachelle Alvarez Rivero 氏の研究は、機械学習の精度を高めるために不完全なデータを補完する方法論を開発する研究である。パターン認識、クラスタリングなどのデータ解析において、解析対象は一つのデータ表現で表されているとは限らない。応用分野によっては、解析対象に対して様々な測定方法があり、複数の観測値の表現が得られることがある。

例えば、タンパク質の機能を予測する問題の場合、ひとつのタンパク質からアミノ酸配列、立体構造、タンパク質間相互作用、発現データなど様々な表現が得られ、それぞれ機能を予測するうえで有用な情報を保持している。アミノ酸配列に相同性があると同じ機能を持ちやすい。アミノ酸配列が折りたたまって構成される立体構造がタンパク質の機能を定める。しばしばほかのタンパク質と相互作用することである機能が成立する。どの条件で発現するかが機能を定める。これらはいずれも機能予測に有用な情報であり、単一のデータ表現を使うよりも、複数の情報を統合したほうが高い予測精度が得られるという研究報告は数多くある。

しかし、それぞれ異なる測定方法から得られることから、しばしばあるタンパク質はいくつかの複数源の情報が得られず、またほかのタンパク質は別の情報源の情報が得られないことがある。従来の方法では一つでもデータ表現が欠けている対象は機械学習に適用できないという問題があった。Rivero 氏は、この問題を解決するため、不完全なデータを補完する 3 つの方法を開発した。

Rivero 氏は、各情報源の情報をカーネル行列と呼ばれる正方行列で表現するアプローチをとった。カーネル行列による表現は、サポートベクトルマシンなど多くの学習機械で用いられている。すると、情報源の個数分のカーネル行列が得られることになる。データの欠損は不完全カーネル行列を生むことになる。すると、不完全なデータを補完する問題は、不完全なカーネル行列どうして補完する問題になる。Rivero 氏はモデルとなる一つの行列を新たに導入し、そのモデル行列を介して、補完問題を最適化問題に帰着させた。目的関数にカルバックライブラ距離と呼ばれる距離尺度を導入することで、情報幾何学の知見を活用することができる。Rivero 氏はこのようにして各ステップが閉形式で表現できるアル

ゴリズムを新たに開発した。

実データを使ったアルゴリズムの検証も行われた。平均値などで単純に不完全行列の欠損要素を埋める方法より、ほかのカーネル行列を相互に利用して補完する方法のほうが、補完精度が高く、かつ、タンパク質の機能予測精度が良くなることを実証した。

以上に述べたように、方法論の新規性、実データを用いた数値実験による実用性の検証も十分と考え、Rivero 氏の学位論文は博士（理工学）の学位に値するものと判定した。

審査年月日 平成30年8月6日

#### 審査委員

主査	群馬大学学術研究院	教授	関 庸一	印
副査	群馬大学学術研究院	教授	太田 直哉	印
副査	群馬大学学術研究院	教授	横尾 英俊	印
副査	群馬大学学術研究院	准教授	横内 寛文	印
副査	群馬大学学術研究院	准教授	加藤 毅	印

#### 関連論文

1 著者名 Rachele Rivero, Richard Lemence, Tsuyoshi Kato

論文題目 Mutual Kernel Matrix Completion (邦訳：相互カーネル行列補完)

雑誌名 IEICE Transactions on Information & Systems 第 E100-D 巻 第 8 号  
1844 頁～1851 頁 2017 年 8 月

※ 掲載決定のものも記載すること