

# Bio-Medical Data Classification Approaches with Limited Annotation

原田, 翔太

<https://hdl.handle.net/2324/4784435>

---

出版情報 : Kyushu University, 2021, 博士 (工学), 課程博士  
バージョン :  
権利関係 :

氏 名	原田 翔太		
論 文 名	Bio-Medical Data Classification Approaches with Limited Annotation (アノテーションが限定された生物・医学データのための分類手法の提案)		
論文調査委員	主 査	九州大学	教授 内田 誠一
	副 査	九州大学	教授 島田 敬士 (システム情報科学府)
	副 査	九州大学	准教授 備瀬 竜馬 (システム情報科学府)

## 論 文 審 査 の 結 果 の 要 旨

医学や生物学の分野で収集される各種データを教師あり機械学習により分類する際、データに十分な教師情報すなわちクラスラベルが付与されていないことが往々にして問題になる。限られた専門家しか正確なクラスラベルを付与できないことがその原因である。一方、深層学習に代表される昨今の機械学習法では、クラスラベル付きデータが不足すると、十分な学習を行うことが不可能となり、結果的に期待した分類精度を達成できない。

この論文は、このクラスラベル付きデータ不足の問題を解決する複数の手法を提案し、実際の医用データを対象とした実験によりそれらの効果を検証することを目的としている。具体的には、対象とする医用データの特性を活かした、データ拡張、グループベースラベリング、および半教師あり学習法をそれぞれ提案している。

第一のデータ拡張では、クラス制約付き敵対的生成ネットワーク (Class-conditional Generative Adversarial Networks, CGAN) に基づく時系列データ生成手法を提案している。それまで主として画像生成用に利用されていた CGAN を時系列用に拡張した同手法により、少数のラベル付きバイタル信号 (EEG や ECG) から各クラスのバイタル信号を人工的に生成することに成功した。それらクラスラベル付き人工データが真のデータに類似していることを確認した後、実際に学習データとして利用することで、深層学習による分類の精度を向上できることを確認した。

第二のグループベースラベリングとは、ラベル無しデータをあらかじめクラスタリングし、各クラスタ内のデータに対して一括して同一ラベルを付与する方法であり、ラベル付与の効率を大幅に向上できる。各クラスタにできる限り同じクラスのデータが含まれるようにするために、一種の制約付きクラスタリングを提案している。制約付きクラスタリングとは、あらかじめいくつかのデータのペアに対して、「同一クラスタに含まれるべき」か「含まれてはいけない」かを指定し、それらを制約としてクラスタリング結果を制御する方法である。これを出発点とし、筆者は内視鏡画像を対象としてその特性を活かした二種類の改良を行っている。第一は制約の緩和である。すなわち、医学的には同一クラスでも画像としては大きく異なるようなペアについては制約を必ずしも満たさなくてよいとした、ソフト制約付きクラスタリング法を実現している。第二は撮影順序の導入であり、連続して撮影された画像は同一クラスタに含まれやすいよう制御している。

第三の半教師あり学習では、クラス分類に重要な特徴だけをデータから抽出する方法として知られる disentanglement 学習 (解きほぐし学習) を、少数のラベル付きデータで実現する方法を提案している。具体的には、内視鏡画像を対象とし、炎症度ラベル付きデータによる disentanglement

学習を進めつつ同時に連続撮影された画像の特徴が近くなるように補正することで、炎症度を表現する特徴の高精度抽出に成功している。

以上を要するに、本論文では、医用データの特性の活用によりラベル付きデータ不足を補うことを可能した機械学習を三手法提案しており、いずれも実データによりその効果を確認している。各手法は、その情報学的新規性に加え、今後様々な実データに応用されうる可能性を持っており、いずれも価値ある業績として認められる。

よって、著者は博士（工学）の学位の資格があるものと認められる。