

令和 3 年 8 月 18 日現在

機関番号：17102

研究種目：挑戦的研究（萌芽）

研究期間：2019～2020

課題番号：19K22895

研究課題名（和文）正例自動サンプリングPU-Learningによるバイオ医療画像解析の省略化

研究課題名（英文）Biomedical image analysis by positive and unlabeled learning

研究代表者

備瀬 竜馬（Bise, Ryoma）

九州大学・システム情報科学研究所・准教授

研究者番号：00644270

交付決定額（研究期間全体）：（直接経費） 4,900,000円

研究成果の概要（和文）：本研究課題では、バイオ医療画像解析特有の対象構造事前知識を用いることで、機械学習に必要な学習データを自動抽出し、アノテーション作業を省力化することを目的とする手法開発を行った。具体的には、対象構造事前知識を活用し正例データを自動抽出し、一部の正例とたくさんのラベル不明サンプルを抽出することで学習可能とするフレームワークを提案し、複数の応用例（病理画像解析、光超音波画像解析）で実証した。これらの研究開発の結果、査読付き論文採択3件（うち2件は、医療画像解析トップ国際会議MICCAIに採択）、投稿2件、学会発表5件、書籍執筆1件、解説記事1件の成果につながった。

研究成果の学術的意義や社会的意義

バイオ医療画像解析分野における認識課題は、疾病の種類や研究目的により様々であること、教師データ作成には専門的な知識が必要であることから、大量の教師あり学習データを作成することが困難である。そこで、本研究では、対象構造事前知識を用いることで、一部の正例を自動抽出し、その他の候補物体をUnlabeledデータとして扱うことで、PU-Learningを適用可能にした。これにより、機械学習に必要な学習データを自動抽出し、アノテーション作業を省力化することが可能となる。光超音波画像や病理画像解析といった複数のタスクに応用し、様々な課題へ展開できることを示した。

研究成果の概要（英文）：In this research, we developed a method that can significantly reduce the annotation in biomedical image analysis by using positive and unlabeled learning, in which our method automatically selects the positive samples using prior knowledge. We also applied this method to several applications such as pathological image analysis and photoacoustic image analysis. As a result, we have published 3 peer-reviewed papers (2 of them were accepted in a top conference in medical image analysis; MICCAI), 2 submissions, 5 conference presentations, 1 book chapter writing.

研究分野：画像情報学

キーワード：機械学習 バイオイメージインフォマティクス 弱教師あり学習 教師なし学習

1. 研究開始当初の背景

バイオ医療画像解析分野において、深層学習を活用した手法が提案されており、成功を収めている。しかしながら、生命科学研究の現場においては、学習を必要としない画像処理ツールである ImageJ 等が一般的に活用されている一方、深層学習の実活用がそれほど進んでいない。これは、一般的な機械学習（教師あり学習）は、十分な量の正例(Positive)と負例(Negative)を用いて学習することを前提としており、そのための教師データを大量に作成することが求められるためだと考える。生命科学研究では、研究目的によって様々な条件下(例: 疾病の種類, 部位, 人種, 細胞種, 成長因子等の細胞培養環境)での実験及び解析が必要となり、画像解析のタスクも多種多様で、様々な研究目的に合わせて個別に専門家が大量の教師データを作成することが高コストであることが大きな要因である。

また、仮にあるデータを用いて教師データを作成して、機械学習モデルを学習したとしても、実際に利用する際は、学習データとテストデータの条件が異なり、機能しない場合がある。例えば、ある設定でせっかく学習データを作成しても、撮像機器や撮像パラメータの違いにより、テストデータでは機能しないという場合が多々あるという課題がある。このような課題に対して、なるべく少ない教師データで学習可能な手法が希求されている。

2. 研究の目的

本研究課題では、バイオ医療画像解析特有の対象構造事前知識を用いることで、機械学習に必要な学習データを自動抽出し、アノテーション作業を省力化することを目的とする。そのために、近年、機械学習分野で盛んに研究がなされている学習手法である Positive and Unlabeled(PU) Learning を用いる。この手法は、一部の正例とたくさんのラベル不明(Unlabeled)サンプルの特徴量空間上の分布から、負例の分布が不明でも、特徴量空間上の適切な識別面を見つけることができるという手法である。

本研究課題では、対象構造事前知識を用いることで、一部の正例を自動抽出し、その他の候補物体を Unlabeled データとして扱うことで、PU-Learning を適用可能とし、適切な識別面を求めることが可能な手法(図1)の研究開発を行う。

PU-Learning 自体が新しい機械学習手法であり、まだバイオ医療画像解析分野には導入されておらず、問題設定自体が新しいチャレンジングな課題といえる。また、本研究課題の特筆すべき点として、テストデータの分布に基づいて学習ができることであり、従来の機械学習の鬼門であったテストデータにおける分布の違いに対応できる点である。本研究課題が達成できれば、煩雑なアノテーションを省略することができ、様々な課題への礎として多面的な応用展開への道を開く。

3. 研究の方法

PU-Learning は、手法が適用する前提条件としていくつかの仮定を置いており、それをどう解決するかが本研究課題を成功させる上で重要なポイント(図1)となる。そこで、下記の2つを考慮しながら、3つの具体的なアプリケーションを対象として手法開発を進めた。

- 正例抽出時及びPU-Learning 時の特徴量空間設計
本研究課題の重要なキーアイデアの一つは、一部の正例抽出時に利用する特徴量空間と PU-Learning 時に利用する特徴量空間を別のものを利用することである。PU-Learning における正例データは、ランダムにサンプリングされていることを前提としており、同じ特徴量空間を利用すると、その仮定に反する。そこで、PU-Learning 時に利用する特徴量とサンプリング時に利用する特徴量がいかに相関のないよう設計できるかがポイントとなる。
- PU-Learning 手法の検討
従来の PU-Learning の前提の一つとして、正例と負例の比率が既知という仮定がある。一般的に、学習データがないと正負の比率は不明なことが多いため、問題ごとに、データにおける正負の比率に依存しない手法検討が必要となる。

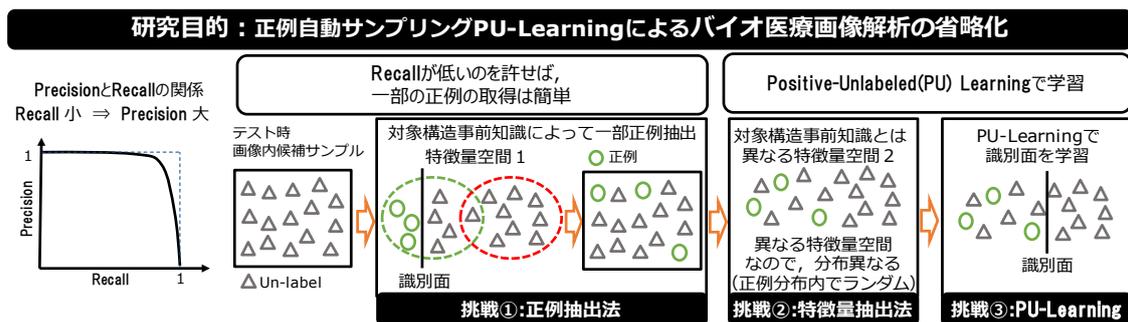


図1. 研究課題概要図

- 1) 光超音波画像における体毛領域の自動認識
光超音波イメージングは、物体が光を吸収して熱膨張することで超音波を発生し、その超音波をセンシングすることで、3次元画像再構成を行う新しいイメージング手法である。血管及び体毛がこの特性を持ち可視化することができる。体毛の数や太さの定量化は美容健康に関する医学研究において重要であり、体毛領域の自動抽出化が望まれている。本研究課題では、体毛領域抽出を当面の対象として研究を進めた。
- 2) 不十分なアノテーションからの細胞検出へのPU Learningの活用
細胞検出用の学習データを作成するためには、画像中の膨大な数の細胞に網羅的に付与する必要がある。そこで、画像中の一部の細胞のみの学習データからでも学習を可能とする手法の開発を行った。
- 3) 病理画像の癌細胞認識へのPU Learningの活用
病理画像において、個々の細胞が腫瘍か正常細胞かを認識することで、病理検体のうち、腫瘍細胞の割合が診断として利用されている。しかし、1枚の病理画像中に数万個の細胞が存在し、個々の細胞にアノテーションすると手間がかかる。そこで、簡便にアノテーションの付与を可能とする新しい手法の開発を行った。

4. 研究成果

- 1) 光超音波画像における体毛領域の自動認識
提案手法では、「体毛は体表に沿って存在し、血管よりも上部にある」という空間構造事前知識を用いて、自動で正例サンプルを抽出し、構造特徴とは経験的に相関が小さい輝度特徴量を利用することで、PU-Learningを適用できるようにした。本研究の利点は、テストデータの分布に基づいて学習できることであり、テストデータにおける分布の違いに対応できる点である。その結果、人によるアノテーション全くなしで、体毛の数や輝度値が全く異なる複数のデータで高い精度 (AUC) を実現した。具体的には、従来の画像処理ベースの手法 0.69 から提案手法 0.91 に精度改善した。
- 2) 不十分なアノテーションからの細胞検出へのPU Learningの活用
この問題では、1枚の画像に細胞が大量に存在する場合に、一部の細胞にしかアノテーションがついていない場合を考える。この不完全な教師データを利用して細胞検出器を学習すると、過検出を行うようになる。この過検出を含む検出点を細胞候補と考え、一部のアノテーションされた細胞を正例サンプルと捉え、それ以外の候補はラベル不明サンプルとして扱うことで、PU Learningを活用可能となる。PU Learningで学習した特徴量を用いて、ラベル不明サンプルの中から尤もらしいサンプルをランキング学習を用いて選択して、細胞検出器を再学習することで、精度を改善していく手法を提案した。複数のデータセットを用いて評価した結果、従来の半教師学習より、平均で 0.14 程度 F-score が改善した。検出結果例を図 2 に示す。
- 3) 病理画像の癌細胞認識へのPU Learningの活用
本手法では、「同一の細胞種が集中して存在」に着目し、領域を囲みその中の細胞全てに一度にラベルを付与することで簡便にアノテーションが可能とした。このとき、正常細胞領域には正常細胞しか含まれないことから正例サンプルとして扱うことができ、がん細胞領域の中には正常細胞も含まれるためラベル不明サンプルとして扱うことで、PU-Learningを適用可能となる。
これらの研究開発の結果、査読付き論文採択 3 件、論文投稿中 2 件、学会発表 5 件、書籍執筆 1 件、解説記事 1 件の成果につながった。査読付き論文のうち 2 本は、医療画像解析分野の世界トップ会議 MICCAI に採択されており、良好な成果を残している。

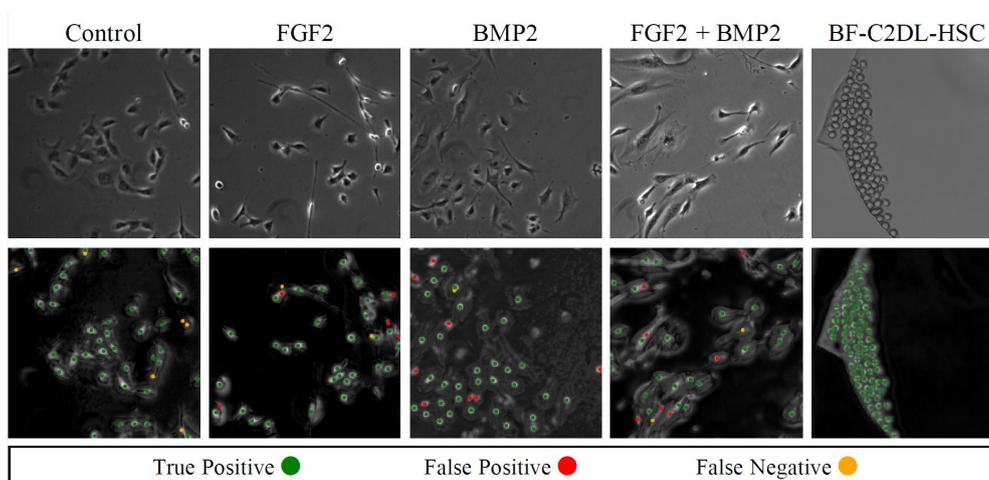


図 2. 提案手法による細胞検出結果例..

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 3件/うち国際共著 1件/うちオープンアクセス 0件）

1. 著者名 R. Bise, K. Abe, H. Hayashi, K. Tanaka, and S. Uchida	4. 巻 5
2. 論文標題 Efficient Soft-Constrained Clustering for Group-Based Labeling	5. 発行年 2019年
3. 雑誌名 MICCAI2019, (top conference in medial image analysis, acceptance rate:31%)	6. 最初と最後の頁 pp.421-430
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/978-3-030-32254-0_47	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 K. Nishimura, E.D. Ker, and R. Bise	4. 巻 1
2. 論文標題 Weakly Supervised Cell Segmentation in Dense by Propagating from Detection Map	5. 発行年 2019年
3. 雑誌名 MICCAI2019, (top conference in medial image analysis, acceptance rate:31%)	6. 最初と最後の頁 pp 649-657
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/978-3-030-32239-7_72	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1. 著者名 K. Nishimura and R. Bise	4. 巻 1
2. 論文標題 Spatial-Temporal Mitosis Detection in Phase-Contrast Microscopy Via Likelihood Map Estimation by 3DCNN	5. 発行年 2020年
3. 雑誌名 Proceedings of 42st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)	6. 最初と最後の頁 pp.1811-1815
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/EMBC44109.2020.9175676	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計5件（うち招待講演 0件/うち国際学会 0件）

1. 発表者名 吉川亮, 備瀬竜馬
2. 発表標題 正例自動サンプリングPositive Unlabeled-Learningを用いた光超音波画像における体毛領域認識
3. 学会等名 パターン認識・メディア理解研究会(PRMU), 2019年9月
4. 発表年 2019年

1. 発表者名 備瀬竜馬, 安部健太郎, 早志英朗, 田中聖人, 内田誠一
2. 発表標題 内視鏡画像のソフト制約クラスタリングによるラベル付け簡略化
3. 学会等名 パターン認識・メディア理解研究会 (PRMU), 2019年9月
4. 発表年 2019年

1. 発表者名 徳永宏樹, 寺本祐記, 吉澤明彦, 備瀬竜馬
2. 発表標題 病理画像癌種別領域分割のための癌種比率を活用した学習手法
3. 学会等名 パターン認識・メディア理解研究会 (PRMU), 2019年12月
4. 発表年 2019年

1. 発表者名 杉本龍彦, 徳永宏樹, Xiaotong Ji, 備瀬竜馬
2. 発表標題 病理画像における陽性細胞の検出
3. 学会等名 電気・情報関係学会九州支部連合大会講演論文集, 2019年9月
4. 発表年 2019年

1. 発表者名 杉本龍彦, 備瀬竜馬
2. 発表標題 PU-Learningを用いた病理画像における簡易アノテーション法の提案
3. 学会等名 画像の認識・理解シンポジウムMIRU2020
4. 発表年 2020年

〔図書〕 計2件

1. 著者名 備瀬竜馬, 内田誠一	4. 発行年 2020年
2. 出版社 羊土社	5. 総ページ数 18
3. 書名 機械学習のさまざまな問題設定と解法(第3章-1-3), 細胞トラッキングの機械学習および数理最適化技術 多物体トラッキングの課題とアプローチ(第3章- -5), 行動追跡とDeepLabCut アニマルポーズトラッキ ング技術(第3章- -6) in 機械学習を生命科学に使う! シークエンスや画像データをどう解析し、新たな 生物学的発見につなげるか?	

1. 著者名 備瀬竜馬, 内田誠一	4. 発行年 2020年
2. 出版社 医歯薬出版株式会社	5. 総ページ数 7
3. 書名 医用画像解析におけるパターン認識, 週間 医学のあゆみ(第五土曜特集「AIが切り拓く未来の医療」)	

〔産業財産権〕

〔その他〕

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------