

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 6 日現在

機関番号：14303

研究種目：若手研究(B)

研究期間：2012～2013

課題番号：24700170

研究課題名(和文) 動的輪郭モデルを用いたグラフベース半教師付き学習に関する研究

研究課題名(英文) Graph-based semi-supervised learning with active contour model

研究代表者

杜 偉薇 (Du, Weiwei)

京都工芸繊維大学・工学科学研究科・助教

研究者番号：00512790

交付決定額(研究期間全体)：(直接経費) 1,800,000円、(間接経費) 540,000円

研究成果の概要(和文)：ラベル情報の少ない学習データを利用して、大量のデータを分類することはパターン認識や機械学習における大きな課題である。実際の利用場面では学習データにノイズデータが混入するため、問題は更に難しくなる。本研究が上記の問題点に基づき、画像処理の分野の一つ手法である動的輪郭モデルで2回微分によってノイズを効果的に削除できる特性を用いて、グラフベース半教師付き学習と組み合わせることによって、データに含まれるノイズを削除し、少ないラベル情報を元にラベルなしデータにラベルを付けることのできる手法を開発した。この手法を人工データと医用画像に応用し、その有効性について示す。

研究成果の概要(英文)：This is a difficult problem on how to classify the unlabeled data in training data and the test data by few labeled data in the field of machine learning and pattern recognition. Especially, some noise data lie on the training data. We propose a graph-based semi-supervised learning with active contour based on the above problems. The algorithm not only removes the noise data, but also classifies the unlabeled data in training data and the test data. The algorithm is applied in the synthetic data and medical images.

研究分野：総合領域

科研費の分科・細目：情報学、知覚情報処理・知能ロボティクス

キーワード：半教師付き学習 パターン認識 医用画像 動的輪郭モデル

1. 研究開始当初の背景

機械学習分野では、近年、大量の学習データに手動でラベルを付ける作業を避けるため、半教師付き学習が盛んに研究されている。半教師付き学習には Self-Training、Co-Training、グラフベースなどの方法が提案されているが、グラフベース手法はデータの構造を図示することができ、アルゴリズムを解釈しやすいという優れた点がある。

われわれはこのグラフベース半教師付き学習をいくつか提案し、画像検索や画像処理に応用してきた。学習データに付与するラベルはできるだけ少なくしたいが、ラベル付けすべき学習データとその個数を定めることは半教師付き学習の非常に難しい問題である。これに対し、まずランダムに選んだ学習データにラベルを一つだけ付ける方法を提案した。しかし、この方法はノイズに敏感であり、結果が不安定になるという欠点があった。そこで、ノイズに対してロバストな方法として、類似度データからファジィクラスタを抽出する半教師付き学習手法を提案した。続いてこの手法を拡張し、ハイパーグラフからの半教師クラスタ抽出法を提案した。更に部分空間の線形写像を非線形写像に拡張し、自然画像とビデオマッピングに適用して認識率が向上することを確認した。これらの手法は画像間色転写にも応用した。その他、データから無用な特徴量を除いたグラフを用いた半教師付き学習手法や、間違っただけラベル付けに対してグラフ上のモードフィルタを用いて誤り訂正をする半教師学習手法を提案してきた。以上の取り組みを通して、グラフベース半教師付き学習においては、たった一つの学習データにラベルを付与だけでも性能を向上させることができることを解明することができた。

以上で述べたように、グラフベース半教師付き学習に関する研究とその応用に取り

組んできた。グラフベース手法では、行列の固有値問題を解いたり、最適化問題を反復法で解く必要があるが、これらの手法はデータにノイズが含まれている場合に敏感や、不安定になるという欠点がある。

この問題の解決が難しい理由は、ノイズの定義が曖昧であることに起因すると申請者は考えた。これに基づいてノイズを特徴ベクトルノイズとクラスラベルノイズに概ね分け、クラスラベルノイズのみにモードフィルタを用いる方法を提案したが、根本的にクラスラベルノイズを削除できなかった。特徴ベクトルノイズが僅かでも含まれると、このアプローチが有効に動かないと考えられる。

一方、画像処理の分野では、動的輪郭モデルでの二階微分によってノイズを効果的に削除できることが知られている。画像のノイズは前述の区分でいう特徴ベクトルノイズと考えられるので、動的輪郭モデルは学習データの特徴ベクトルノイズの除去にも有効ではないかと予想される。これが動的輪郭モデルを用いたグラフベース半教師付き学習という着想に至った理由である。

本研究では動的輪郭モデルをクラスラベルノイズにも拡張利用しようと考えており、医用画像データなどの実データを使って検証したい。

2. 研究目的

ラベル情報の少ない学習データを利用して、大量のデータを分類することはパターン認識や機械学習における大きな課題である。実際の利用場面では学習データにノイズが混入するため、問題は更に難しくなる。

本研究ではグラフベース半教師付き学習と動的輪郭のモデルを組み合わせることで、データに含まれるノイズを削除し、少ないラベル情報を元にラベルなしデータにラベルを付けることのできる新しい手法を提案する。また、提案手法を医用画像の

ベンチマークデータに適用し、その結果をROC 曲線(受信者動作特性曲線)で評価する。

本研究で提案した半教師付き学習手法がわずかなラベルしか付与されておらず、且つ、ノイズを含むような学習データであっても、それを利用して大量のデータを分類することができる。これを画像処理分野の手法である動的輪郭モデルを学習データからのノイズ除去手法として利用する点である。動的輪郭モデルは、画像処理において輪郭を抽出するための手法であるが、ノイズの影響を受けにくいという長所を持つ。よって、ノイズの混入した学習データからノイズを削除することにも有効である。

3. 研究の方法

動的輪郭モデルを用いた半教師付き学習手法を提案・開発し、学習データに含まれたノイズを削除する。また、ラベルの付与されていないデータに正しいラベルを付ける。医用画像のベンチマークデータに対して本手法を応用して評価する。

表 1: 図 1 の説明

class1:	○ラベルなしデータ
	●クラスノイズ
	▲ラベル付けたデータ
class2:	●ラベルなしデータ
	○クラスノイズ
	■ラベル付けたデータ
	X特徴ベクトルノイズ

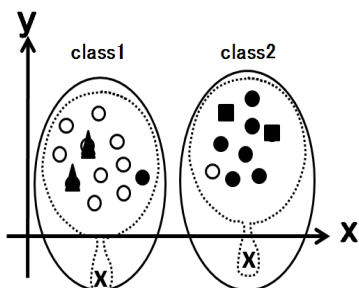


図 1: ノイズを含んだ学習データ

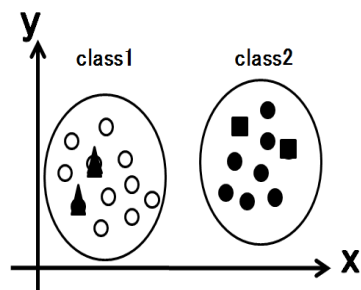


図 2: ノイズを削除した学習データ

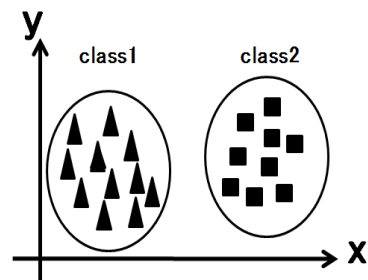


図 3: 分類した学習データ

図 1 にノイズを含んだ学習データの図を示す。表 1 に記号の説明を示す。ここでクラス 1 とクラス 2 を分類したい。各クラスには、ラベルなしデータ、ラベルありデータ、特徴ベクトルノイズ、クラスラベルノイズが含まれる。動的輪郭モデルで特徴ベクトルノイズを削除することができる(図 2)。また、動的輪郭モデルを用いたグラフベース半教師付き学習の手法で図 2、図 3 のようなクラスラベルノイズを削除することができる上でラベルなしデータに正しいラベルを付ける。

動的輪郭モデルは、内部エネルギーと外部エネルギーの和を最小にするように振る舞うモデルである。このモデルの初期値とパラメータの設定は難点である。初期値に対しては、3 つの情况进行を検討した。(1). 動的輪郭モデルの初期値をノイズではない、学習データのラベルとする。ラベルなしデータを参考しなくても、ノイズを除去できた。(2). (1) と同様に動的輪郭モデルの初期値をノイズではない、学習データのラベルとし、今度はラベルなしデータを参考しながら、ノイズを除去できた。(3). 動的輪郭モデルの初期値をランダムに与え、ノイズを除去するため、ラベルなしデータを参照できるた。パラメータ

に対しては、各初期値の設定毎に、パラメータを調整して最適な範囲を求めた。定式化されたグラフ上の関数回帰によって半教師付き学習のラベル伝搬手法のエネルギーの2次関数を動的輪郭モデルのエネルギー関数に変えて反復法で解を求めた。

4. 研究成果

- (1). 学習データに含まれるノイズの形成原因とその定義(ノイズの分類)を明らかにした。
- (2). ノイズを含む学習データの学習への影響を明らかにした。またノイズ(誤ったラベル)が存在している状況で、ラベルなしデータが学習に対して有効な意味を持つかどうかを検証した。
- (3). 動的輪郭モデル用いた新しいグラフベース半教師付き学習を開発し、これによって除去できるノイズの種類を明らかにした。
- (4). 提案した手法を医用画像に応用するため、MRI(Magnetic Resonance Imaging)の基本原則と応用を調査した。

本研究で提案する半教師付き学習手法がわずかなラベルしか付与されておらず、且つ、ノイズを含むような学習データであっても、それを利用して大量のデータを分類することができる。これを画像処理分野の手法である動的輪郭モデルを使って行うという点に学術的な新規性と意義がある。また、闇雲にノイズ除去を行うのではなく、ノイズを分類してから除去するというアプローチも広めていきたい。

5. 主な発表論文等

[学会発表](計2件)

- (1). W. Du, Hybrid DWI Gradient Strength and Separation Effects in Prostate Cancer Diagnosis, Joint Annual Meeting ISMRM-ESMRMB 2014, May 2014, Milan, Italy
- (2). W. Du, Object Segmentation by Edges Features of Graph Cuts, the Seventh International Conference on Software Engineering Advances 2012, Nov. 2012,

Lisbon, Portugal

6. 研究組織

(1) 研究代表者

杜 偉薇 (Du, Weiwei)

京都工芸繊維大学・工芸科学研究科・助教

研究者番号：00512790