

# 11章 11.2まで

## 11.0 冒頭文

### 11.1 はじめに

前提条件1 データが少ない

そのための主な手法

前提条件2 ラベルなしデータ

それ以外の手法

### 11.2 データ拡張

#### 11.2.1 基本的な考え方

前提条件 データ拡張

#### 11.2.2 主に画像に対して有効な方法

## 11.0 冒頭文

1. 本書で紹介している手法は、多くの訓練データが必要という制限がある
  - a. これを用意するのは難しい
  - b. 少数データで分析する方法を模索する
2. 以下の手法で、そのような制限下の分析を可能にする
  - a. データ拡張
  - b. 転移学習
  - c. 半教師あり学習
  - d. 自己教師学習
  - e. マルチタスク学習
  - f. ドメイン適応
  - g. 少数事例学習
  - h. 能動学習

## 11.1 はじめに

### 前提条件1 データが少ない

1. 手元のデータからタスク $\mathcal{T}$ のために学習したい
  - a. 主にクラス分類を想定する
  - b. しかし、大部分はクラス分類以外でも使える
2. 手元のデータは $\mathcal{D} = \{(\mathbf{x}_n, d_n)\}_{n=1, \dots, N}$ 
  - a.  $N$ は小さい
  - b. だが、これを用いて学習用の準備を行う

## そのための主な手法

1. データ拡張(11.2節)
  - a. 後述の手法(前述のc.からh.の手法)のもとになる
  - b. 単体でもよくつかわれる
2. 転移学習(11.3節)
  - a. 広い意味ではタスク $\mathcal{T}'$ の学習経験を利用することを指す
    - i. 日経平均株価の予測で使ったモデルをダウ平均株価の予測に活かす、のような話？
  - b. 狭い意味だと特定の手法を指す
    - i. ここでは具体的な言及なし

## 前提条件2 ラベルなしデータ

1. 手元のデータは $\mathcal{D}_{\text{UL}} = \{(\mathbf{x}_n)\}_{n=1, \dots, N}$ 
  - a. 正解ラベルがない
  - b. 半教師あり学習や自己教師あり学習は、このようなデータを対象としている

## それ以外の手法

その他の手法の概要は、以下の通り。

マルチタスク学習	異なる複数タスクを1つのネットワークで学習する
ドメイン適応	訓練データとテストデータの間で統計的な分布にずれがある場合に適用する
少数事例学習	訓練データの数が極端に小さい場合を対象にした学習

## 11.2 データ拡張

### 11.2.1 基本的な考え方

#### 前提条件 データ拡張

1.  $\mathcal{D} = \{(\mathbf{x}_n, d_n)\}_{n=1, \dots, N}$  をもっている
  - a.  $\mathbf{x}_n$  を適当な変換  $\varphi$  で  $\mathbf{x}'_n \equiv \varphi(\mathbf{x}_n)$  とする
  - b. 新たなサンプル  $(\mathbf{x}'_n, d_n)$  として利用する
2. 変換は自由度が高いが、以下の制約などがある
  - a. 変換によってラベルが変わらない
  - b. テストデータに出てきそうなデータにしないといけない
3. 数だけは際限なく増やすことができる

### 11.2.2 主に画像に対して有効な方法

- そもそもデータ拡張は画像データに対して有用
- 画像データを変化させても正解ラベルが変化しないような変換をする必要がある
- 図11.1にあるような変換を加える