第4章

単語集合間の関連度計算手法

4.1　序言

本章では，前章で述べた時空間データのマルチモーダル学習モデルの学習手法について述べる．まず，??節で一般的なディープネットの学習の流れを説明する．次に，??節降水量予測モデルの学習で用いる最適化の手法について述べる．最後に，??節で本章のまとめを示す．

4.2　ディープネットの学習手法

ディープネットの学習手法として，バッチ学習とオンライン学習がある．本節では，前者のバッチ学習について説明を行う．まず，??節でバッチ学習の手法について述べる．次に，??節でバッチ学習の一種であるミニバッチ学習の手法について述べる．

4.2.1　バッチ学習

ニューラルネットへの入力 $\boldsymbol{x}$ と対応する目標 $\boldsymbol{y}$ の組から成る訓練データ

$$\mathcal{D} = \{(\boldsymbol{x}_1, \boldsymbol{y}_1), ..., (\boldsymbol{x}_N, \boldsymbol{y}_N)\} \tag{4.1}$$

が与えられ，訓練データの入力を $X = (\boldsymbol{x}_1, ..., \boldsymbol{x}_N)$，目標を $Y = (\boldsymbol{y}_1, ..., \boldsymbol{y}_N)$ とする．ニューラルネットの学習は，訓練データをもとに計算される誤差関数 $E(X, Y; \boldsymbol{\theta})$ を，ネットワークのパラメータ $\boldsymbol{\theta} = [W_1, \boldsymbol{b}_1, ..., W_M, \boldsymbol{b}_M]$($M$ は $\boldsymbol{\theta}$ の成分数) について最小化することである．なお，訓練データの全サンプルによる誤差 $E(X, Y; \boldsymbol{\theta})$ は，一つのサンプルに対する誤差関数 $E_n(\boldsymbol{x}_n, \boldsymbol{y}_n; \boldsymbol{\theta})$ を用いて次のように計算される．

$$E(X, Y; \boldsymbol{\theta}) = \sum_{n=1}^{N} E_n(\boldsymbol{x}_n, \boldsymbol{y}_n; \boldsymbol{\theta}) \tag{4.2}$$

一般に，ニューラルネットの学習は誤差逆伝播を行い，**勾配降下法 (Gradient Descent Method)** によりネットワークのパラメータを逐次更新していくことでなされる．すなわち，式 (??) で表される誤差関数 $E(X, Y; \boldsymbol{\theta})$ の $\boldsymbol{\theta}$ による微分 (勾配)

$$\nabla E = \frac{\partial E}{\partial \boldsymbol{\theta}} = \left[\frac{\partial E}{\partial W_1} \cdots \frac{\partial E}{\partial \boldsymbol{b}_M} \right]^{\mathrm{T}} \tag{4.3}$$

をもとに，現在のネットワークのパラメータ $\boldsymbol{\theta}$ を負の勾配方向 $(-\nabla E)$ に少しずつ動かすことで，パラメータの逐次更新を行う．従って，勾配降下法における i $(i = 1, ...)$ 番目のイテレーションでは，パラメータ $\boldsymbol{\theta}$ の更新は式 (??) に基づいて行われる．

$$\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i - \alpha \nabla E \tag{4.4}$$

式 (??) でのパラメータの更新について，更新する大きさ (勾配 ∇E に掛けられる係数)α は**学習率 (Learning rate)** と呼ばれ，学習がうまく行えるかどうかを左右する重要なハイパーパラメータである．勾配降下法では，ネットワークのパラメータの初期値 $\boldsymbol{\theta}_1$ を適当に決め，式 (??) に示した更新式によってパラメータ $\boldsymbol{\theta}_2$,$\boldsymbol{\theta}_3$,... を得る．従って勾配降下法のアルゴリズムは次のようになる．ただし，**Algorithm**

Algorithm 1 勾配降下法

1: initialize $\boldsymbol{\theta}_1$, $i \leftarrow 1$
2: **for** *epoch* $\leftarrow 1$ to MAX_EPOCHS **do**
3: $\quad \boldsymbol{\theta}_{i+1} \leftarrow \boldsymbol{\theta}_i - \alpha \nabla E(X, Y; \boldsymbol{\theta}_i)$
4: $\quad i \leftarrow i + 1$

??中の値 MAX_EPOCHS は，最大学習回数を表す定数である．**Algorithm ??**でのバッチ学習では，訓練データ $\mathcal{D}$ 全てについての誤差 $E(X, Y; \boldsymbol{\theta})$ を計算し，誤差関数の勾配 ∇E を用いて学習を行っている．訓練データ全てをもとにパラメータを更新する学習手法を**バッチ学習**と呼ぶ.

4.2.2 ミニバッチ学習

??節で述べたバッチ学習では，全ての訓練データをもとにパラメータを更新していた．一方で，訓練データの一部だけを用いてパラメータを更新する学習手法として**ミニバッチ学習**がある．ミニバッチ学習では，訓練データ $\mathcal{D}$ を，複数の訓練データ $\mathcal{D}_1$,..., $\mathcal{D}_M$ に分割し，各訓練データ $\mathcal{D}_m$ についての誤差 $E(X_m, Y_m; \boldsymbol{\theta})$ を式 (??) によって求める．

$$E(X_m, Y_m; \boldsymbol{\theta}) = \frac{1}{N_m} \sum_{(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{D}_m} E_n(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\theta}) \tag{4.5}$$

なお，式 (??) 中の値 N_m はミニバッチ $\mathcal{D}_m$ に含まれるサンプルの数を表す．ミニバッチ $\mathcal{D}_m$ に対する誤差 E を求めたのち，誤差関数の勾配 ∇E を用いて式 (??) に基づいてパラメータの更新を行う．

$$\begin{gathered}\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i - \alpha \nabla E_m \\ \text{where } E_m = E(X_m, Y_m; \boldsymbol{\theta_i})\end{gathered} \tag{4.6}$$

式 (??) を用いたミニバッチ学習の手法は**確率的勾配降下法 (Stochastic Gradient Descent; SGD)** と呼ばれており，アルゴリズムは次のようになる．なお，

Algorithm 2 確率的勾配降下法

initialize $\boldsymbol{\theta}_1$, $i \leftarrow 1$
for *epoch* $\leftarrow 1$ to MAX_EPOCHS **do**
  for $m \leftarrow 1$ to M **do**
    $\boldsymbol{\theta}_{i+1} \leftarrow \boldsymbol{\theta}_i - \alpha \nabla E(X_m, Y_m; \boldsymbol{\theta}_i)$
    $i \leftarrow i + 1$

Algorithm ??中の値 M は，ミニバッチ $\mathcal{D}_1, \ldots, \mathcal{D}_M$ の個数を指す．

4.3 降水量予測モデルの学習に用いる最適化手法

降水量予測モデルの実装において，ディープネットの学習ではミニバッチ学習を行い，最適化手法としてモメンタム付き確率的勾配降下法，RMSProp を用いる．本節では，モメンタム付き確率的勾配降下法，および RMSProp ついてそれぞれ??節，??節で述べる．

4.3.1　モメンタム付き確率的勾配降下法

ディープネットのミニバッチ学習では，しばしば??節で述べた確率的勾配降下法が用いられる．確率的勾配降下法の派生として，パラメータ更新の勢いを表すモメンタム (Momentum) を導入し，学習性能を向上させたモメンタム付き確率的勾配降下法がある．モメンタム付き確率的勾配降下法では，各イテレーションにおいてパラメータの更新量 ∇E_m を保存しておく変数 $\boldsymbol{v}_i$ を導入した，式 (??) が用いられる．

$$\begin{gathered} \boldsymbol{v}_i = \gamma \boldsymbol{v}_{i-1} + \alpha \nabla E_m \\ \boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i - \boldsymbol{v}_i \\ \text{where } E_m = E(X_m, Y_m; \boldsymbol{\theta_i}) \end{gathered} \tag{4.7}$$

ただし，式 (??) 中の γ は，学習において現在の勾配の更新方向をどれだけ保つかを決定するハイパーパラメータである．モメンタムを導入した確率的勾配降下法のアルゴリズムは次のようになる．

Algorithm 3 モメンタム付き確率的勾配降下法

```
initialize θ_1, v_0 ← 0, i ← 1
for epoch ← 1 to MAX_EPOCHS do
    for m ← 1 to M do
        v_i ← γ v_{i−1} + α∇E(X_m, Y_m; θ_i)
        θ_{i+1} ← θ_i − v_i
        i ← i + 1
```

4.3.2　RMSProp

??節で述べた確率的勾配降下法と同じく，ミニバッチ学習を行う際の最適化手法として **RMSProp** がある [?]．RMSProp は，各イテレーションでのパラメータ

の更新に式 (??) を用いる．

$$
\begin{aligned}
\boldsymbol{v}_i &= \gamma \boldsymbol{v}_{i-1} + (1-\gamma)\left(\nabla E_m\right)^2 \\
\boldsymbol{\theta}_{i+1} &= \boldsymbol{\theta}_i - \alpha \frac{\nabla E_m}{\sqrt{\boldsymbol{v}_i + \epsilon}} \\
\text{where } E_m &= E(X_m, Y_m; \boldsymbol{\theta_i})
\end{aligned}
\tag{4.8}
$$

ただし，式 (??) 中の γ は**減衰率 (Decay Rate)** と呼ばれるハイパーパラメータ，α は学習率である．RMSProp のアルゴリズムは次のようになる．

Algorithm 4 RMSProp 最適化法

```
1: initialize θ_1, v_0 ← 0, i ← 1
2: for epoch ← 1 to MAX_EPOCHS do
3:     for m ← 1 to M do
4:         δ_i ← ∇E(X_m, Y_m; θ_i)
5:         v_i ← γ v_{i-1} + (1 − γ) δ_i^2
6:         θ_{i+1} ← θ_i − α δ / sqrt(v_i + ε)
7:         i ← i + 1
```

4.4 結言

本章では，前章までで説明した時空間データのマルチモーダル学習モデルの学習手法について説明を行った．ディープネットの学習手法として，バッチ学習とオンライン学習を挙げ，特にバッチ学習について説明を行った．また，降水量予測モデルの実装において用いた 2 種類の最適化手法，すなわちモメンタム付き確率的勾配降下法と RMSProp について説明を行った．