東京大学 量子ソフトウェア寄付講座

第6回 量子ソフトウェアハンズオン 2024年 7月 29日

本日の演習の概要

- □ ニューラルネットワークのモデルをテンソルネットワーク技術を活用して圧縮する技術について、学びます。
- □ 演習を通じて、以下の目標を達成することを目指します。

本日の目標

- 1. ニューラルネットワークのモデルから全結合層を取り出し、行列積演算子に変換することで、 パラメータ数を削減する方法を学ぶ。
- 2. 圧縮されたモデルの性能を確認し、パラメータ数と性能のトレードオフを確認する。
- 3. 圧縮された状態のモデルでの追加学習や再学習を試み、学習の様子を確認する。

■ 演習の内容

- □ 前述の目標を達成するため、本日は以下の演習に取り組んでいただきます。
- □ 分からない点などがあれば、積極的に周囲の人と相談しながら取り組んでください。
 - □ 講師側でも適宜サポートに回ります。

| # | 演習テーマ | 概要 |
|---|------------------------------|---|
| 1 | データセットの確認とニューラルネッ トワークの学習 | 今回利用するデータセットについて内容を確認します。ベースラインとなるニューラルネットワークモデルの作成を行います。 |
| 2 | 行列積演算子を用いたニューラル ネットワークの分解 | ニューラルネットワークの全結合層を取り出し、MPOへ変換する手順を確認します。 MPOをニューラルネットワークの中に埋め込み、モデルとしての性能を確認します。 パラメータ数と変換の精度の関連について確認します。 |
| 3 | グループ演習 | ニューラルネットワーク、テンソルネットワーク、学習に関する様々なハイパー パラメータを調整しながら、モデルのサイズ、性能などの変化を確認してい ただきます。 |

▋■演習環境の準備

- 今回の演習はGoogle Colaboratoryを使って、演習・解説を行います。
- 各自のPCのローカル環境上で行っていただいても構いませんが、ご自身の環境に依存するエラー対応については、自己責任でお願いいたします。

今回の講義のgithubページに以下のノートブックを掲載してあります。

• 「01_行列積演算子を用いたニューラルネットワークの分解.ipynb」

ご自身のドライブにコピーしていただき、Google Colab環境でノートブックを開いてください。

- □ 今回用いるデータは、とある金融機関における金融商品の電話営業の実績データです。
- □ このデータを元に、商品を購入してくれそうなユーザを推測するニューラルネットワークモデルを作成します。
 - 電話営業の効率化を行うようなタスクをイメージしていただければと思います。

・ 利用できる情報の種類

ユーザの属性情報:年齢、職業、学歴など

電話営業に関する情報:電話をした月/曜日、通話回数、通話時間など

社会情勢に関する情報:消費者物価指数、金利など

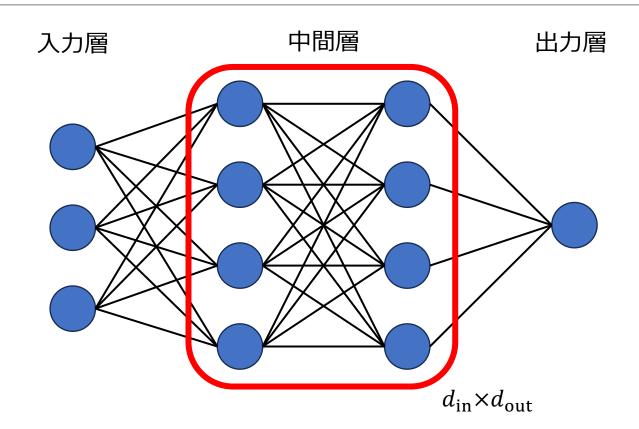
・ ターゲットとなる情報

最終的に商品を購入したかどうか(Yes / No)

データの出典

Moro,S., Rita,P., and Cortez,P.. (2012). Bank Marketing. UCI Machine Learning Repository. https://doi.org/10.24432/C5K306.

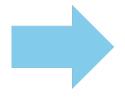
- □ ニューラルネットワークは、入力ベクトルと出力ベクトルの関係を数理的に表現するためのモデルです。
- □ モデルのパラメータを更新することで、入力ベクトルと出力ベクトルの関係を近似するモデルを作ります。
- □ 全結合層は入力次元×出力次元の行列を用いて線形変換を行う層である。一般に、この後に活性 化関数と呼ばれる非線形変換を行い、表現力を高めます。
- 層数、ニューロン数を増やすことでモデルの表現力が高まるが、過学習のリスクも高くなります。



- □ ニューラルネットワークには直接文字列の情報を入力することができないので、これを数値に変換します。
- □ ここでは、One-Hot Encodingと呼ばれる手法を用いて数値に変換します。
- □ また、その他の数値情報を含め、すべて0~1の範囲に正規化しておきます。

One-Hot Encodingのイメージ図

| | 年代 |
|-----|-----|
| Aさん | 20代 |
| Bさん | 30代 |
| Cさん | 20代 |
| Dさん | 40代 |
| Eさん | 40代 |



| | 20代 | 30代 | 40代 |
|-----|-----|-----|-----|
| Aさん | 1 | 0 | 0 |
| Bさん | 0 | 1 | 0 |
| Cさん | 1 | 0 | 0 |
| Dさん | 0 | 0 | 1 |
| Eさん | 0 | 0 | 1 |

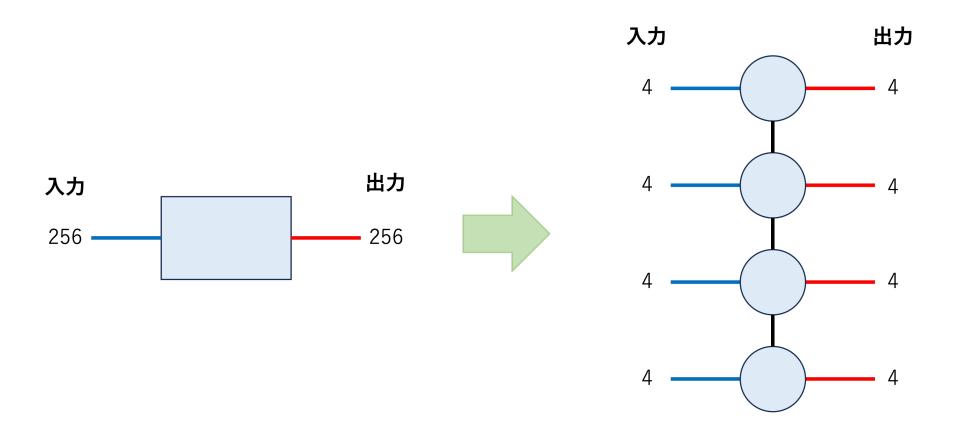
- □ ニューラルネットワークの学習が完了したら、性能を確認します。
- □ 今回は正例が10%程度の不均衡データを用いるので、何を評価指標とするかの設計は実用上は 重要なのですが、ここでは、利用が簡単なROC-AUCを中心に見ていくことにします。

各指標の簡単な説明

| 指標名 | 概要 | |
|----------------|---|--|
| Accuracy (正答率) | 正しく予測できたデータの割合 | |
| Recall(再現率) | 実際の正例のうち、正しく正例と予測できた割合 | |
| Precision(精度) | 正例と予測したもののうち、実際に正例である割合 | |
| ROC-AUC | スコアの並び順がどの程度適切かを評価する指標 (上位に正例がどのぐらい集まっているか) | |

■ 演習2. 行列積演算子を用いたニューラルネットワークの分解

- □ ニューラルネットワークの全結合層は入力次元×出力次元の行列であり、中間層は大きな行列を用いることも多いです。(数千×数千ぐらいのオーダー。)
- □ この部分を取り出し、より少ないパラメータ数で、再現するために、テンソルネットワークを活用します。



■グループ演習

- □ ここまでの演習を通じて、ニューラルネットワークのモデルを、テンソルネットワークの技術を活用して圧縮する方法について学びました。
- □ グループ演習では、様々なパラメータを変更してもらいながら、性能とパラメータのトレードオフや、効率的な活用について、実験・考察していただきます。
- □ 元のモデルの学習、追加学習、圧縮後のモデルでの再学習など、様々なパターンから興味のあるテーマを探して、実験を行っていただき、グループ内で共有してください。

テーマの一例 (これ以外の内容を実施していただいても問題ございません。)

| # | テーマ案 | 概要 |
|---|----------------------|---|
| 1 | 行列積演算子のボンド次元を変化させる | モデルの圧縮率と精度のトレードオフを確認してみます。また、追加学習や再学習に与える影響を確認します。 |
| 2 | 元のネットワーク構造を変化させる | より大きな/小さなモデルを作り、これを圧縮してみることで、ベストなモデルサイズを探ります。 |
| 3 | 学習のハイパーパラメータを変更する | オプティマイザや学習率を変化させて、影響をみます。特に、圧縮後のモデルでの影響を確認します。 |
| 4 | (発展)様々な分解の形を試す | 行列積演算子への変換の形は一意ではないため、様々な 形の分解を試して、性能への影響を確認します。 |
| 5 | (発展)他のモデル圧縮手法と比較してみる | ・ 枝刈り・量子化などの他のモデル軽量化手法との比較をします。 |

全体ディスカッション

- □ 各グループでメンバーが取り組んだ内容について、整理して、全体に向けて発表していただきます。
- □ 今回の演習を通じて得られた知見や、活用に向けたビジョン・課題感などを共有してください。