

画風変換LoRAの内部パラメータによるモデルの埋め込み表現の獲得

金田 悠路 (静岡大学)

背景 LoRAの共有が当たり前

様々な業務をLoRAによって管理

- ロゴLoRA → ブランドロゴ製作
- 画像生成AI + キャラLoRA → ブランドキャラ製作
- 画像生成AI + 図解LoRA → マニュアル製作

LoRAを共有するプラットフォームの存在

CIVITAI

水墨画風LoRA ジブリ風LoRA アニメ風LoRA

数十万件のLoRAが公開中

様々なタスクに特化させたLoRAを大量に保有・運用

ニーズ LoRAモデルをベクトルとして扱いたい!

メリット① LoRAモデル間の... 距離計算が可能!

ベクトル空間内でどちらが近い?

類似度比較

クラスタリング

LoRAモデルの特徴を考慮した比較が可能に

メリット② LoRAモデルを... 機械学習で分析可能!

機械学習モデル

ベクトルに変換し入力可能!

分類

ランキング

内部パラメータによるLoRAモデル検索も実現可能に

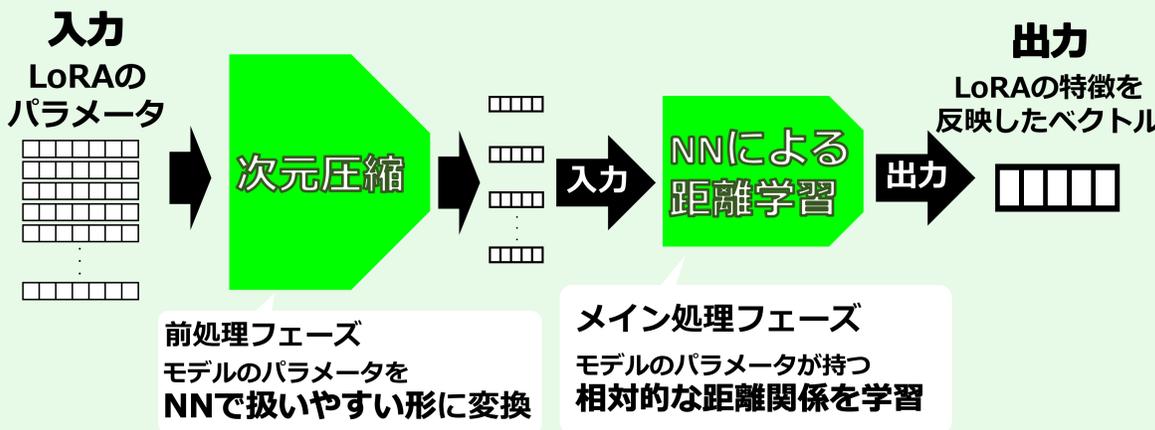
メタデータでのベクトル化が一般的だが...

- メタデータ
- 出力例

知らないLoRAをベクトル化したい!

LoRAモデルの内部パラメータからモデルをただか数次元の密ベクトルに!

提案 LoRAのパラメータから次元圧縮 × 距離学習によってベクトル化!



これを実現する3つの工夫で、**構造的情報**・**パラメータ**を考慮したベクトル化

工夫① パラメータのFlat化

LoRAのsafetensorsファイル

異なるランクのLoRAを同じ次元で扱う工夫

レイヤごとに低ランク圧縮行列を掛け合わせて元の行列を復元

レイヤごとにFlat化

LoRAのFlat化ベクトル系列

- パラメータ特徴
- 構造的情報

を保持しつつ共通の形式に

工夫② PCAによる次元圧縮

LoRA1のFlat化したベクトル系列

LoRA2のFlat化したベクトル系列

レイヤごとに次元圧縮

Incremental PCA

レイヤごとに特徴量を抽出

LoRA1の次元圧縮ベクトル系列

LoRA2の次元圧縮ベクトル系列

LoRAのFlat化したベクトル系列

各レイヤの特徴を保持したレイヤ数分の次元圧縮系列に構成

工夫③ Triplet Net型のTransformer Encoderによる距離学習

LoRA1, LoRA2, LoRA3

提案エンコーダ

絶対位置エンコーディング

Multi Head Attention

FFN

MLP

Triplet Loss

レイヤ間の相互関係を考慮

絶対位置エンコーディングの追加で構造的情報を付与

レイヤ間の関連度を学習

構造的な重要性を考慮

LoRAの各レイヤが持つ情報量の差がある

各レイヤの重みを算出

加重平均を出力

各レイヤの構造的な重要性を学習

相対的な特徴の学習

Triplet Lossで学習後の潜在空間

似ているペアを近くに、似ていないペアを遠くに学習

Anchorに対してその観点における相対的な類似性を学習

実験①: 推論性能に関するアブレーションテスト

知見①: 位置埋め込み・MLPによる加重平均の組み合わせが埋め込み表現の質を向上

アブレーションテストの結果

手法名	Triplet Loss (エラー率)	Triplet Acc (正解率)
提案手法	0.136	0.871
MLPなし	0.147	0.856
位置情報なし	0.208	0.779
ベースライン	0.157	0.832
距離学習無し	0.418	0.545

距離学習により構造的類似性を学習

提案手法が最も高い結果に (統計的に有意)

二つの工夫を併用することで各レイヤの構造的情報に基づく重要度に応じた集約が可能に

実験②: 埋め込み表現と人間の類似性判断との一致度

知見②: 学習時の正解としたタグの類似性が変換画像の類似性と不一致

LoRAの変換画像の類似性をLoRAの類似性として被験者がラベル付け

人手でラベル付けしたTripletsに対する推論評価結果 (各手法は大規模データセットで学習)

手法名	Triplet Loss	Triplet Accuracy
提案手法	0.432	0.500
MLPなし	0.426	0.527
位置情報なし	0.414	0.507
ベースライン	0.408	0.569

全体的に低精度

提案手法がほかの手法と比べてやや劣る結果に (統計的に有意差なし)

学習時の正解としたLoRAモデルにつくタグの類似性

実験時の正解としたLoRAモデルの変換特徴の類似性

実験③: 埋め込み表現を用いたモデル検索性能

知見③: 検索タスクにおいてTriplet Lossによる局所的順序関係の学習では不十分

作成したクエリに類似するLoRAランキングを推論したLoRAベクトルの類似度から作成

クエリ

順位	LoRA
1	ステッカー
2	キャラ
3	図解

変換画像をもとにしたランキング性能結果 (各手法は大規模データセットで学習)

手法名	MAP@10	nDCG@10	MRR
提案手法	0.463	0.850	0.581
MLPなし	0.442	0.850	0.517
位置情報なし	0.429	0.868	0.508
ベースライン	0.434	0.847	0.490

どの指標において手法間で大きな差なし

nDCG@10以外で提案手法が最も高い精度 (統計的に有意差なし)

三つ組の相対的な類似性を学習するTriplet Lossは全体的な順序の最適化が不十分

LoRAの重みのみから特徴を抽出する枠組みが技術的に可能に!!