8/1 Discussion

実験結果と論文調査

TR24006 小林 亮太（中部大学工学部ロボット理工学科　藤吉研究室）

http://mprg.jp

MPRG
MACHINE PERCEPTION AND ROBOTICS GROUP

はじめに

- 追加実験
  - 概要
  - 結果
- 今後のテーマ

追加実験

実験概要

- MCNを用いて近づけるモーダルの組み合わせ方による学習効果を調査
  - 3モーダルを二段階で学習する実験：V（ビデオ），A（オーディオ），T（テキスト）
    - AV_T：AとVで学習してからTを追加
    - VT_A：VとTで学習してからAを追加
    - AT_V：AとTで学習してからVを追加

- 卒論時は低い精度
  - エポック数が不足
  - 追加の実験が必要

実験結果

- データセットによる精度の違い
  - YouCook2　：AT_Vが最高精度　卒論時と同様
  - MSR-VTT　：VT_Aが最高精度

	YouCook2			MSR-VTT		
	R@1	R@5	R@10	R@1	R@5	R@10
AV_T	1.26	4.18	6.96	2.69	8.06	14.05
VT_A	6.84	16.2	22.6	6.20	14.5	20.1
AT_V	8.06	17.3	21.6	1.65	5.27	8.37

評価用データセットの性質とモーダル組み合わせが
紐づいて精度に影響

実験結果　卒論との比較　YouCook2

- 追加実験前後での精度の違い

YouCook2						
	卒論			追加実験		
	R@1	R@5	R@10	R@1	R@5	R@10
AV_T	1.61	5.97	9.43	1.26	4.18	6.96
VT_A	1.10	5.13	7.13	6.84	16.2	22.6
AT_V	5.16	11.3	15.3	8.06	17.3	21.6

VT_AとAT__Vは精度が大幅に増加→エポック不足
AV_Tは若干の低下

実験結果　卒論との比較　MSR-VTT

- 追加実験前後での精度の違い

MSR-VTT						
	卒論			追加実験		
	R@1	R@5	R@10	R@1	R@5	R@10
AV_T	0.18	0.92	1.63	2.69	8.06	14.05
VT_A	0.14	0.53	1.14	6.20	14.5	20.1
AT_V	0.14	0.61	1.35	1.65	5.27	8.37

全てのパターンで大幅に増加
卒論時の実験で設定ミスしていた可能性

実験結果　再現実験との比較

- 追加実験と再現実験での精度の違い

		YouCook			MSR−VTT		
		R@1	R@5	R@10	R@1	R@5	R@10
追加実験	AV_T	1.26	4.18	6.96	2.69	8.06	14.05
	VT_A	6.84	16.2	22.6	6.20	14.5	20.1
	AT_V	8.06	17.3	21.6	1.65	5.27	8.37
再現実験							

全てのパターンで大幅に増加
卒論時の実験で設定ミスしていた可能性

現在の実験後の予定

- 未定

- 当初のテーマである知識転移グラフに立ち戻る

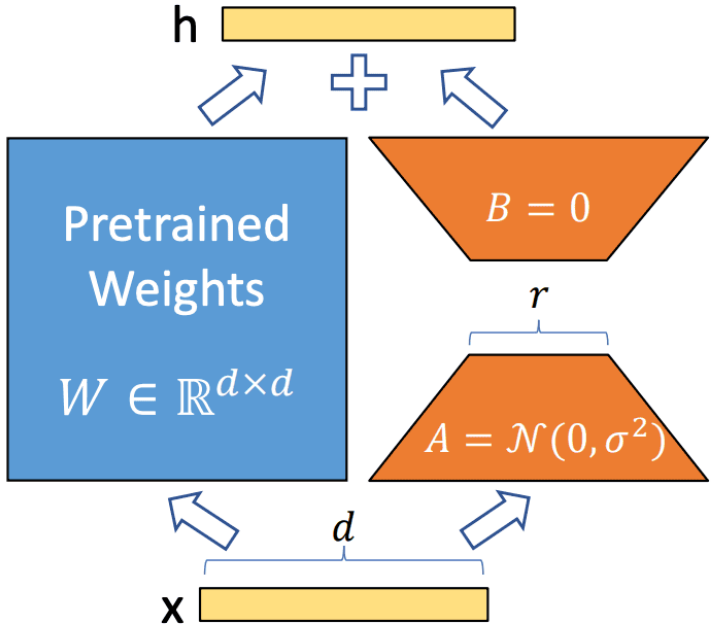
- 対照学習時にLoRAのような形でくっつけて「LoRAを介して特定のモーダル間を近づけることに特化した学習」をモーダルの組み合わせごとに同時に行って，共通の特徴量の学習と下流タスクに応じてLoRAを使い分けることで各下流タスクで高い精度を発揮可能な仕組みみたいなこと

論文調査

LoRa：Low-Rank Adaptation of Large Language Models[Edward Hu+, ICLR'22]

- 背景
  - 大規模言語モデルの進化
  - ファインチューニングの課題に対処
- LoRaの提案
  - 事前学習済みモデルの重みの固定
  - 低ランク分解行列を挿入

$$W + AB$$



LoRa：Low-Rank Adaptation of Large Language Models[Edward Hu+, ICLR'22]

- LoRaの提案

$$W + AB$$

$$\begin{pmatrix} 2 & 3 & 1 & \cdots \\ 6 & 8 & 0 & \cdots \\ 9 & 7 & 4 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

(m × n)
固定

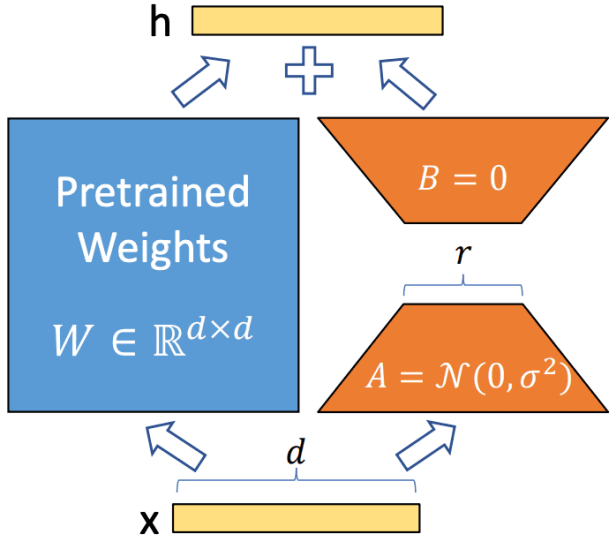
$$\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ \vdots \end{pmatrix}$$

(m × k)

$$(1 \quad 0 \quad 1 \quad 0 \quad \cdots)$$

(k × n)

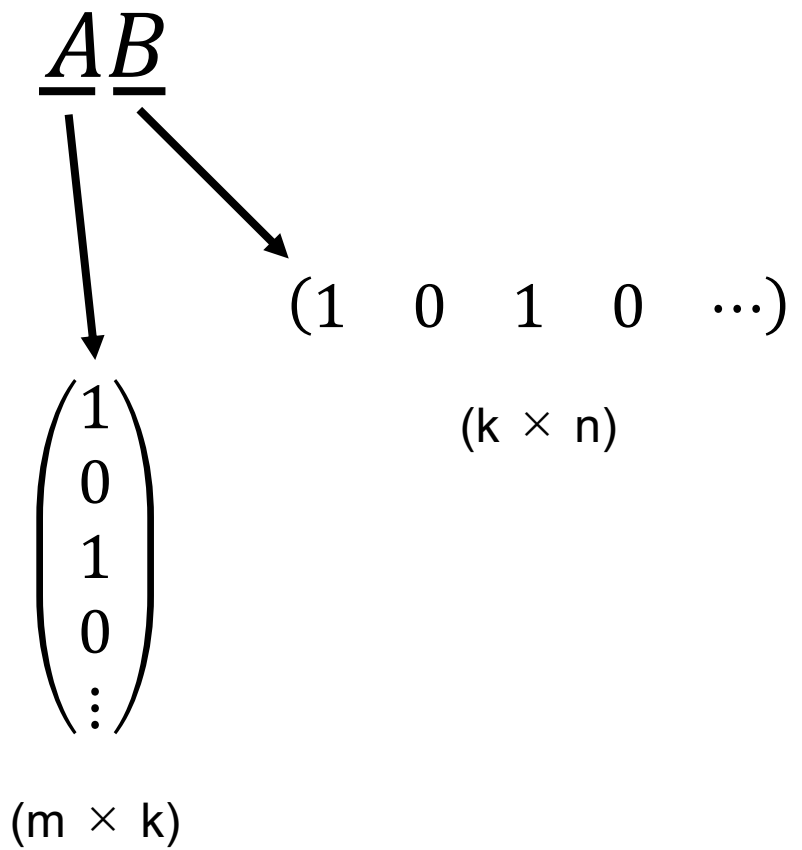
(m × n)



LoRAの利点

- AとBのパラメータだけを更新

- パラメータ数を大幅に削減
- メモリ効率の向上
  - 最大で1/3のVRAM使用量



今後の予定

おわりに

- 実験　　：完了　卒論時の実験の続き

- 今後の予定：
  - 今後の方針を固める

資料

Improving Discriminative Multi-Modal Learning with Large-Scale Pre-Trained Models [C Du, ICLR'24]

- 大規模の事前学習モデルを活用したマルチモーダル学習の改善
- LoRAを各モーダルのエンコーダに適用
  - 一部または全てのモーダル