

グラスマン多様体上における fewshot learning について

中島 秀斗 統計的機械学習研究センター 特任助教

1 概要

現在の機械学習は非常に多くのデータを必要としている。一方で、人間は数少ないデータから学習することができる。Fewshot learningは、人間が如何にして学習しているかについて研究するメタラーニングの一種で、学習済みのモデルを用いて、学習では用いなかったクラスについて判別を試みるものである。また、多様体は数学的に扱いやすい構造を持った集合であり、より複雑な構造を持つデータに対応するために多様体を用いた機械学習も広く研究がなされている。ここでは、グラスマン多様体上における fewshot learning についての研究を紹介する。本研究は福水健次氏との共同研究である。

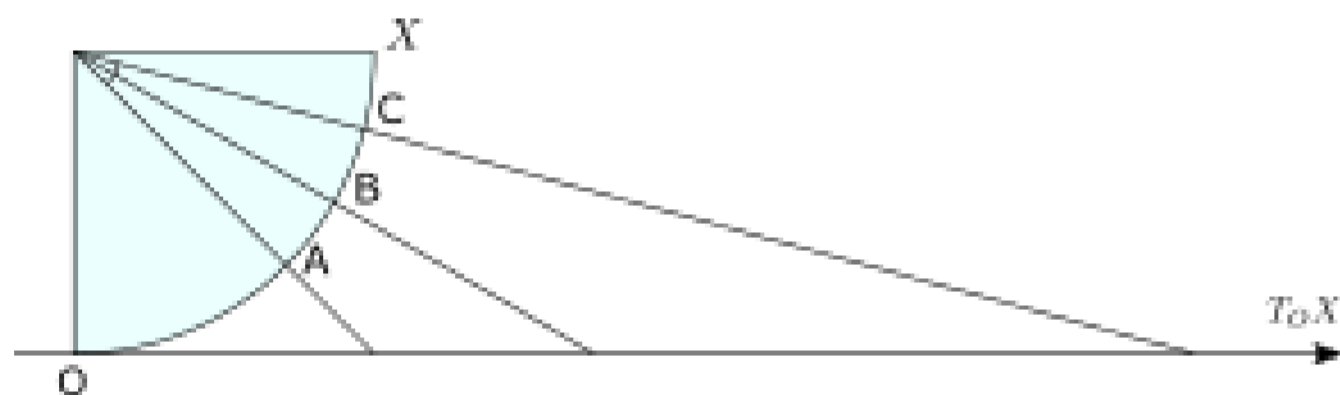
2 グラスマン多様体

グラスマン多様体 $Gr_N(k)$ は「数ベクトル空間 \mathbb{R}^N 内の k 次元部分空間」を要素とする多様体である。これは $N \times k$ 行列のなす空間において、正則行列の右から作用に関する同値類で割ることにより得られる商空間を、幾何学的な言葉で表現したものである。

また、グラスマン多様体には自然な群作用（正則行列を左から掛ける作用）を持つ等質空間でもあり、接平面や測地線距離などが具体的に計算できる。このことから、球面や双曲空間などと並んで等質空間の典型例と見做されている。

測地距離と接平面上での距離について

下図のように多様体 X 上の3点A,B,Cにおいて、ABおよびBCの測地線距離が等しい場合であっても、対応する接平面上での距離は等しいとは限らない。



グラスマン多様体の場合は、基準点から離れば離れるほど、その差異は大きくなっていく。すべての点を等価に扱う必要がある場合はこの性質は不利であるが、ある一点を特別視して考える場合などは、この性質が有利に働くことが期待できる。

3 目的関数

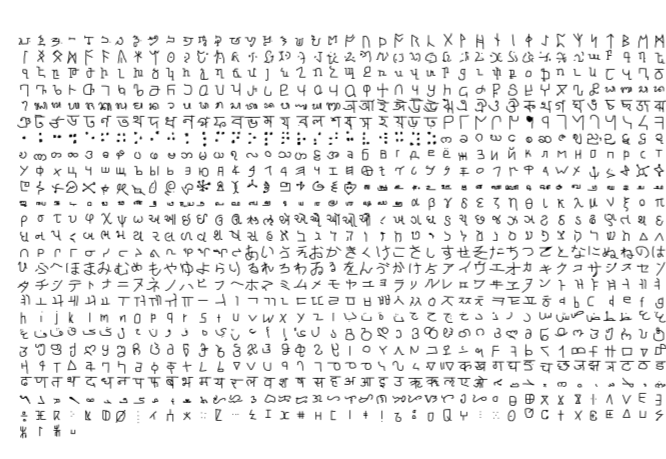
判別分析のアイデアを利用するとともに、各クラスの代表元(templates)を作成することにより、計算コスト軽減を図る。具体的には、以下の目的関数を最小化する。

$$L(X; U, W) = - \sum_{a=1}^N \sum_{i=1}^K \frac{\exp\{-d(WX_i^{(a)}, WU^{(a)})\}}{\sum_{b=1}^N \exp\{-d(WX_i^{(a)}, WU^{(b)})\}}$$

ただし、 $X_i^{(a)}$ はクラス $a \in \{1, \dots, N\}$ に属するデータ、 $U^{(a)}$ はクラス a の代表元、 W は正則行列である。また d は距離関数で、接空間上の距離 d_{tangent} (基準点は第2引数) または測地線距離 d_{geod} のいずれかである。

Omniglotについて

Omniglot は世界各地の言語で用いられている文字(架空文字含む)からなるデータセットで、30言語(全964文字)からなるbackground、20言語(全659文字)からなるevaluationで構成され、文字だけで見れば1623のクラスがある。各クラスには20個ずつの画像が用意されている。各クラス毎のデータが少ないため、fewshot learning 等でよく用いられている。



4 問題設定

メタラーニングにおいてよく用いられている手法 N -way K -shot を用いる。すなわち、 N 種類のクラスからそれぞれ K 個のデータが与えられたとき、これらのみの情報から N 種類のクラスの要素を判別できるようにしたい。右図は4-way 3-shotの例である。

Support set	Queries
あ あ あ	あ
い い い	い
う う う	う
え え え	え

そのために、データセットからランダムに N 個のクラスを選び、そこから K 個のデータを取り出して学習用データを作成する。そして、選んだ N 個のクラスから学習に用いなかったデータを利用して N 択問題を Q 個作成してその正答率を見る。以上を1タスクとして、それを T 回繰り返して、 N 択問題 (全 $T \times N$ 問) の正答率を調べる。

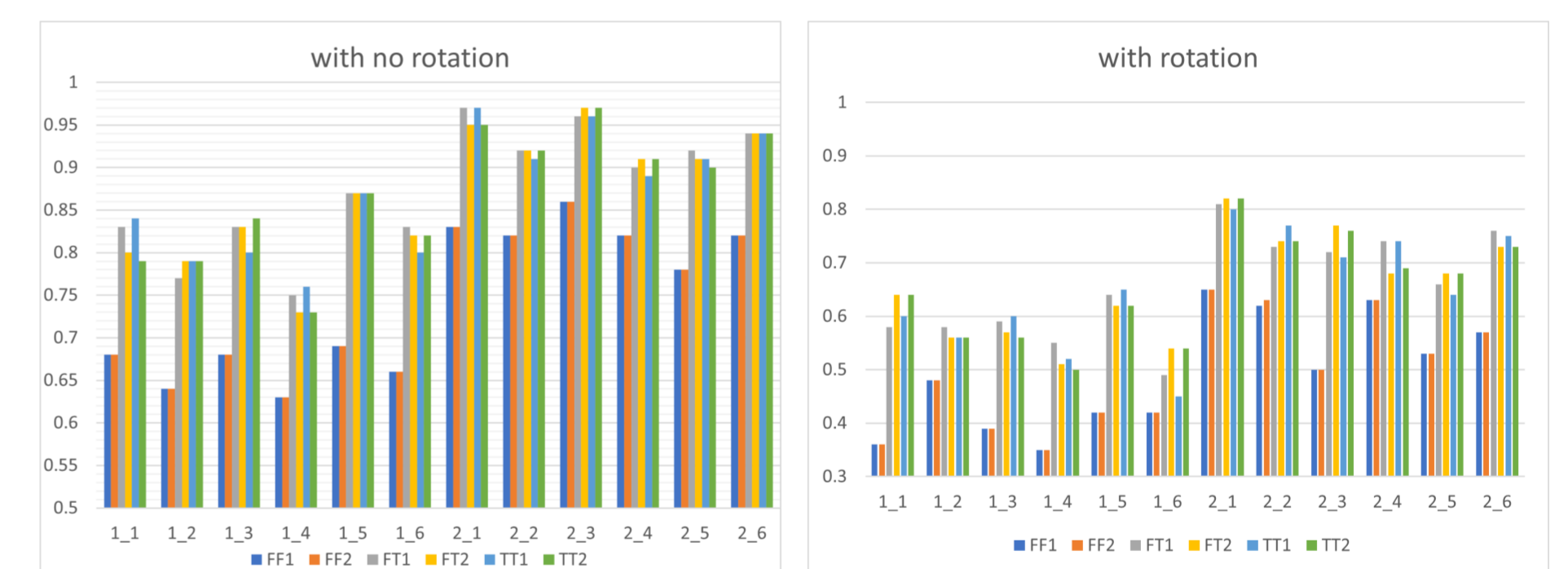
5 実験

データセット Omniglot に属するデータを ResNet18 を用いて行列に変換し、それをグラスマン多様体の要素として扱う。データの複雑性を加味するため、画像の回転(90° 毎)の有無いずれの場合でも実験を行っている。本実験では $(N, K) = (5, 5)$ の5-way 5-shot であり、 $Q = 5, T = 20$ としている。

Omniglot からデータを生成する際に、2系統を考察した。一つは同一言語の文字から5クラスを選ぶ系統(1_n)、もう一つはすべての言語の文字から5クラスを選ぶ系統(2_n)である。

実験に関するグラフ

2種類の距離に対して、(a) 学習しない場合(1,2), (b) W のみ学習する場合(3,4), (c) U と W いずれも学習する場合(5,6), についてそれぞれ実験を行った。ここで括弧内の数字は、下図のグラフにおける左側から数えた時の位置を表しており、奇数が接空間上の距離 d_{tangent} に、偶数が測地線距離 d_{geod} に対応している。また左図は画像の回転をしない場合、右図は回転を含めた場合での実験結果である。



代表元 $U^{(a)}$ の初期値は $X_i^{(a)}$ の平均を取ったもの、 W の初期値は、単位行列をノルムに関して正規化したものを取る。エポック数は各タスク毎に300。また W に関しては、ノルムが1になるように正規化を行う。問題設定上、選ばれるクラスに応じて問題難易度が大きく変わる。設定による差を比較するために、結果の平均等の統計処理を行っていない。

6 考察・結論

実験のサンプル数や各実験でのデータ数が十分ではないためこの実験データから結論を出すことはできないが、群作用により性能は大きく改善される一方で、代表元の学習は大きく影響しない傾向にあることが分かった。また、距離の変更に関しては顕著な差は見られなかった。余談であるが特徴を抽出して得られる行列をグラスマン多様体の要素としてみたとき、同一言語間の文字の方が、全言語間の文字に比べて判別が難しい傾向にあるようである。この傾向は文字情報から言語の推定などへの応用が期待される。

参考文献

[1] Ting-Yao Hu, Zhi-Qi Cheng, and Alexander G. Hauptmann, Subspace representation learning for few-shot image classification, arXiv:2105.00379v2.