

論文読みまとめ

Meta-learning for Few-shot Natural Language Processing: A Survey

井上智裕

2021 年 1 月 20 日

1 導入

1.1 扱う問題

メタ学習、Few-shot natural language processing

1.2 問題意識

今回対象とする問題は、サンプルデータサイズが小さいタスクである。こういったタスクは 1 クラスあたりのデータ数を k として k -shot learning (one-shot learning, few-shot learning など) と呼ばれる。従来のニューラルネット (NN) は、ラベル付けされたデータセットに対し、反復学習を行うことで学習を行う。NN でモデルの精度を向上させるためにはモデルのサイズを大きくしたり、多数のデータセットを時間をかけて学習させたりすることが有効である。しかし、データの少ないタスクではこの方法を取ることはできない。そこで、データの少ないタスクを迅速に学習できる効率的なアルゴリズムの構築を目的とする方法としてメタ学習を紹介する。

2 メタ学習

2.1 メタ学習とは

メタ学習とは、学習方法を学習するためのアルゴリズムである。ニューラルネット (NN) では、ある特定のタスクにおいて、ラベル付けされた入力例を用いて学習を行う。図 2.1[1] に NN の学習プロセスを示す。まず、入力例に対するモデルの予測ラベルと正解ラベルから loss 関数により loss を計算する。次に、この loss が小さくなるように optimizer を用いてネットワークのパラメタを調整する。このように、NN では train サンプルと test サンプルが同じ分布に由来するものであるという仮定のもと、train サンプルを用いて、test サンプルにも対応できるようにネットワークのパラメタを学習する。

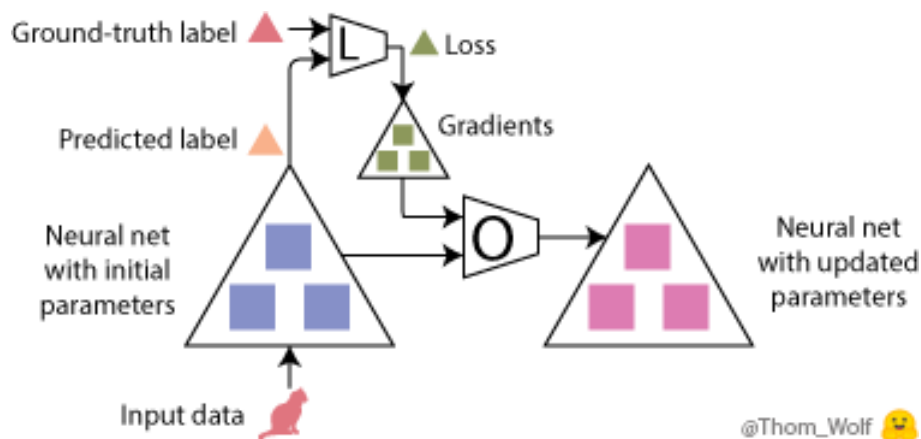


図 2.1 NN の学習プロセス [1]

一方、メタ学習では、特定のタスクを解決の対象とするのではなく、タスクの集合に対して、タスク 1 つを学習サンプル 1 つとして扱い、学習プロセス全体を学習する。図 2.2[1] にメタ学習の学習プロセスを示す。まず、あるタスク T_i のデータセットに対して、通常の NN 同様に学習を行う。学習されたモデルのタスク T_i における test loss が meta loss として機能する。次に、この meta loss が小さくなるように meta optimizer により meta parameter の学習が行われる。meta parameter としては、NN の初期パラメタや optimizer のパラメタが用いられる。このように、メタ学習では train タスクと test タスクが同じ分布に由来するものであるという仮定のもと、train タスクを用いて、test タスクを効率よく学習できるように meta parameter を学習する。

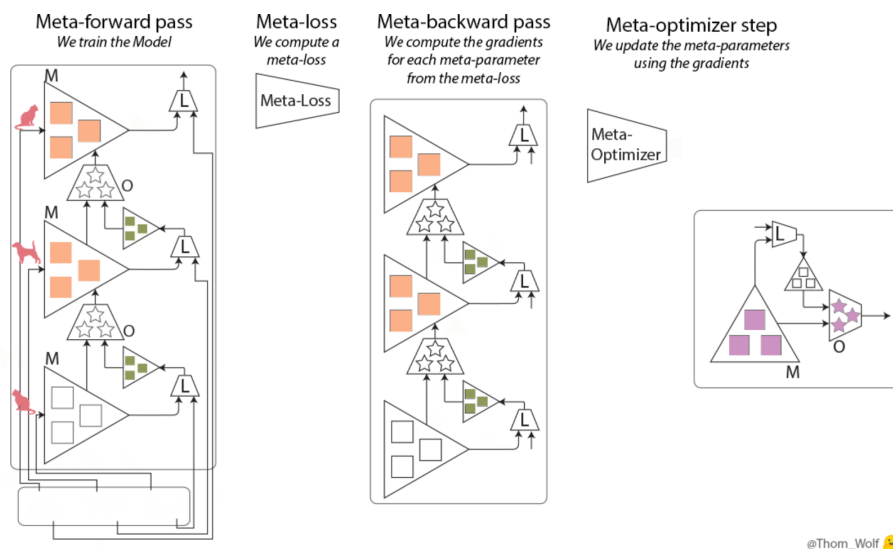


図 2.2 メタ学習の学習プロセス [1]

2.2 メタ学習と転移学習の比較

転移学習は、事前にあるタスク A のデータセットを用いてネットワークのパラメタを事前学習させておき、ターゲットタスクであるタスク B のデータセットで追加の学習を行ってモデルの調整を行うという逐次的な学習手法である。転移学習はタスク A の学習により同じ問題領域にあるタスク B に対応できるという学習方法であるのに対して、メタ学習は効率の良い学習方法を学習することができるので適応できるタスクはさらに広い。人間の学習方法に例えると、転移学習は数学を学ぶと理科の成績が良くなることに似ていて、メタ学習は勉強の際にいつの演習まで復習するかといった方法論自体を学ぶことで全く異なる領域の勉強にも応用が利くことに似ている。

2.3 メタ学習とマルチタスク学習の比較

マルチタスク学習は、複数のタスクに対応できる単一のモデルを構築する手法のことで、関連する複数のタスクを同時に学習させることで実現する。複数のタスクに依存する点でメタ学習はマルチタスク学習の一種であるといえるが、これらには異なる点が存在する。まず、マルチタスク学習はターゲットタスクに対する汎化性能の高い事前学習モデルを学習することが目的である。一方で、メタ学習はターゲットタスクを素早く学習する効率的な学習アルゴリズムを学習しようとするのが目的である。また、マルチタスク学習では、データ数の多いタスクほど有利になるためデータ数が主な関心となるのに対して、メタ学習では、学習タスクの数が多いほど有利になるため学習タスク数が主な関心となる。

3 メタ学習の歴史

3.1 埋め込みの学習：metric-based meta-learning

metric learning はデータ間の距離関数を学習する手法のことである。直感的には意味の近いデータは近く、意味の遠いデータは遠くなるように計量を学習する [2]。距離関数はインスタンスを表現空間に埋め込む関数と空間内での類似度を計算する関数（余弦類似度やユークリッド距離）からなる。距離関数が学習タスクで十分に学習されていれば、目標タスクでは fine-tuning なしで十分に機能する頑健性の高い手法である。

3.1.1 マハラノビス距離学習

古典的な metric learning の手法 [2]。式 3.1 の共分散行列 M を学習する。

$$d_M(x, y) = \sqrt{(x - y)^T M (x - y)} \quad (3.1)$$

M が半正定値行列であればマハラノビス距離は以下のように変形できる。

$$\begin{aligned} d_M(x, y) &= \sqrt{(x - y)^T M (x - y)} \\ &= \sqrt{(x - y)^T L^T L (x - y)} \\ &= \sqrt{\{L(x - y)\}^T L(x - y)} \\ &= \|Lx - Ly\|_2 \end{aligned} \quad (3.2)$$

これは、 x, y を実正方行列 L により表現空間に埋め込み、そこでのユークリッド距離を取っていると見なすことができる。なお、 M は非類似サンプルの組の集合 D と類似サンプルの組の集合 S に対して、以下の制約

付き最適化問題を解けばよい。

$$\begin{aligned}
 M &= \arg \max_M \sum_{(x_i, x_j) \in D} d_M(x_i, x_j) \\
 \text{subject to } &\sum_{(x_i, x_j) \in S} d_M^2(x_i, x_j) \leq 1, \quad M \geq 0
 \end{aligned} \tag{3.3}$$

3.1.2 Siamese Network

Siamese Network は、2006 年ごろに提案された手法である。その概要を図 3.1 に示す。Siamese Network は、2 つのインスタンスを入力として受け取り、同じクラスに属す場合は 1、そうでない場合は 0 を出力する。埋め込み関数 f には deep learning を用いることができ、こうした deep learning を導入した metric learning を deep metric learning という。

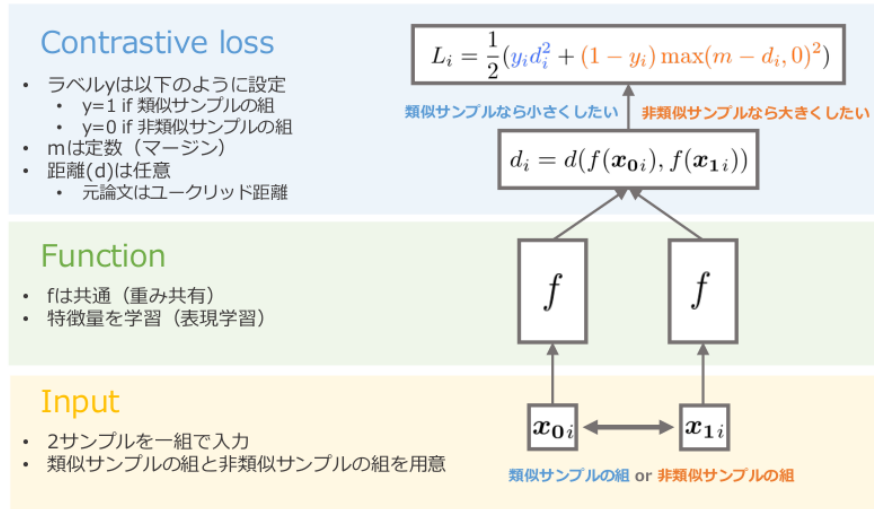


図 3.1 Siamese Network の概要 [2]

3.1.3 Matching Network

Matching Network は、2016 年に提案された one-shot 分類問題に対する最初の metric-based なメタ学習アルゴリズムである。Matching Network は、パラメトリックな最近傍探索アルゴリズムが基本となっており、式 3.4 で定義される。

$$P(\hat{y}|\hat{x}, S) = \sum_{i=1}^k a(\hat{x}, x_i) y_i \tag{3.4}$$

S は各クラス 1 サンプルずつサイズ k のラベル付きデータセットである。また、 a は NN により得られる \hat{x}, x_i の表現に対して余弦類似度を計算する関数である。Siamese Network が本質的には埋め込み空間における距離関数であったのに対して、Matching Network は埋め込み空間内におけるデータセット S を用いた最近傍分類器と見なすことができる。Matching Network では、表現の埋め込みがデータセット S 全体の影響を受けているという特徴がある。

3.1.4 Prototypical Network

Prototypical Network は、2017 年に提案された few-shot 分類問題に対する手法で Matching Network と比較して、2 つの新しい点がある。1 つ目に、類似度計算にユークリッド距離を用いることである。2 つ目に、データセットの各サンプルの埋め込み表現に対する距離を利用するのではなく、図 3.2 に示すようにクラスごとの埋め込み表現の平均をプロトタイプ c_k としてプロトタイプ c_k との距離を利用することである。

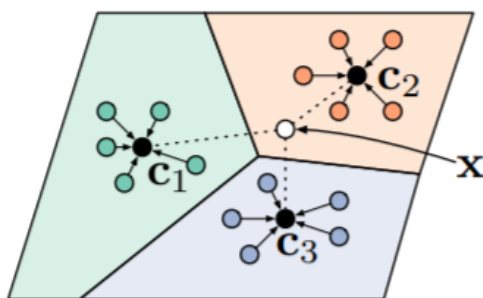


図 3.2 Prototypical Network の距離計算 [3]

3.1.5 Relation Network

Relation Network は、2018 年に提案された手法であり、metric を次式で定義する。

$$r_{i,j} = s(e(x_i), e(x_j)) \quad (3.5)$$

ここで、関数 e は入力インスタンスの表現ベクトルを生成する埋め込み関数である。また、 s は類似度を計算する関数であり、Prototypical Network とは異なり、DNN である。本質的にここまであげた metric learning は事前学習された最近傍探索アルゴリズムである。

参考文献

- [1] Thomas Wolf. From zero to research — an introduction to meta-learning — by thomas wolf — huggingface — medium. (Accessed on 01/19/2021).
- [2] Metric learning 入門 - cypaste の日記. (Accessed on 01/19/2021).
- [3] Jake Snell, Kevin Swersky, and Richard S. Zemel. Prototypical networks for few-shot learning, 2017.