

無限に大きな機械学習モデルの理論的分析

加納龍一, 杉山磨人

(総合研究大学院大学, 国立情報学研究所)

どんな研究?

巨大化を続ける機械学習モデルの性質を、数学の力を使い明らかにしています。

数十、数百パラメータだったのに、コンピュータの進化で最近は何兆に! ?

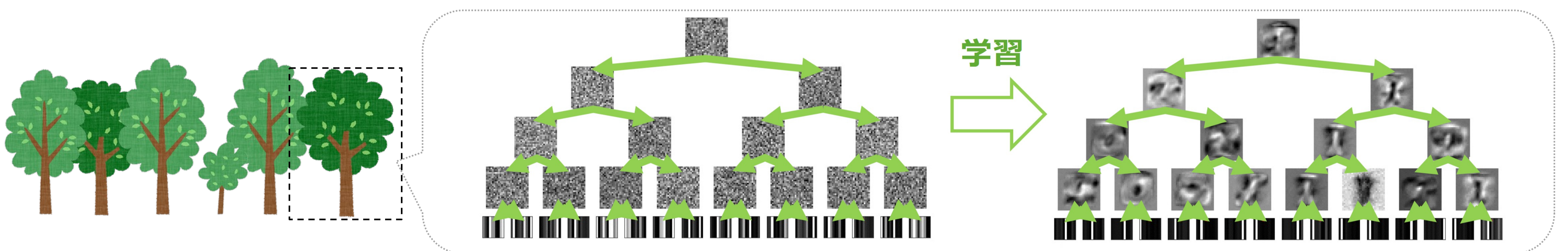
何がわかる?

巨大モデルの精度が良い理論的な理由は、世界的にまだよく理解できていません。理解できると、巨大モデルを使う上での安心感に繋がります。

顔認証、自動運転、...

背景・目的

大量の木から出力を得るモデル(木アンサンブル)を考えます。深層学習と並ぶモデルで、実社会でも活躍しています。我々は、木の本数を無限大にしたときの挙動を理論解析しました。



研究内容 (方法・結果・結論)

方法

無限の木が誘発するニューラルタンジェントカーネルを導出し、訓練挙動をカーネル法の枠組みで記述します。

パラメータ空間の勾配法

$$\frac{\partial \theta_\tau}{\partial \tau} = -\eta \frac{\partial L(\theta_\tau)}{\partial \theta_\tau} = -\eta \frac{\partial f(\mathbf{x}, \theta_\tau)}{\partial \theta_\tau} (f(\mathbf{x}, \theta_\tau) - \mathbf{y})$$

関数空間の勾配法

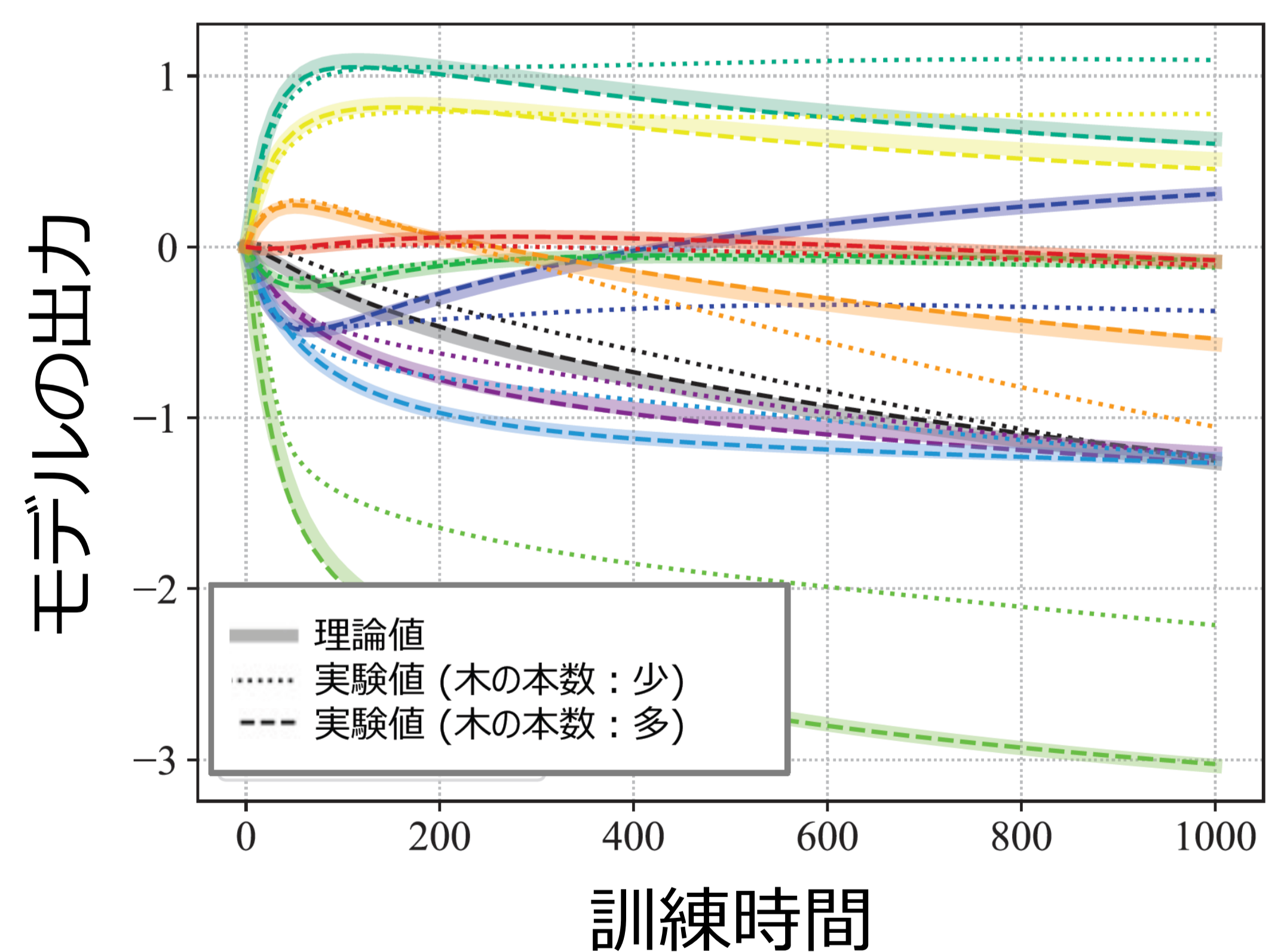
$$\begin{aligned} \frac{\partial f(\mathbf{x}, \theta_\tau)}{\partial \tau} &= \frac{\partial f(\mathbf{x}, \theta_\tau)}{\partial \theta_\tau} \frac{\partial \theta_\tau}{\partial \tau} \\ &= -\eta \underbrace{\frac{\partial f(\mathbf{x}, \theta_\tau)^\top}{\partial \theta_\tau} \frac{\partial f(\mathbf{x}, \theta_\tau)}{\partial \theta_\tau}}_{\text{Neural Tangent Kernel}} (f(\mathbf{x}, \theta_\tau) - \mathbf{y}) \end{aligned}$$

我々は、ニューラルタンジェントカーネルが持つ特性を理論的に調査しました。

その結果、**無限本の木の挙動は解析的に記述可能**となることがわかりました。

結果

ニューラルタンジェントカーネルを使用し、無限本の木の訓練挙動を解析的に記述すること成功しました。



また、この取り組みは様々な新しい発見に繋がりました。

- 無限に大きくなっても、過度に性能は劣化しない
- 深さごとの葉の枚数が同じであれば挙動は一致する
- 木の中でのパラメータ共有により性能は変化しない

