

グラフスペクトルを介した深層グラフモデルの漸近挙動解析

東京大学, Preferred Networks 大野健太
東京大学, 理研 AIP 鈴木大慈

本講演ではグラフスペクトルを介したグラフ CNN の表現能力に関する講演者らの研究を報告する。グラフ CNN はグラフ構造データの解析に利用される深層学習モデルであり、化合物・ソーシャルネットワーク・ポリゴン画像など様々な分野に応用されている。一般的な深層モデルでは層（変換の基本単位）を増やし、さらに層間に非線形関数を挟むことで、表現能力が向上する（すなわち、グラフ CNN が表現できる関数全体の空間が大きくなる）ことが理論的・経験的に示されている。一方、グラフ CNN においてはしばしば層を増やしても予測精度が向上しない現象が実験的に報告されている。この問題の理論的解明のため、本研究ではグラフ CNN の層数無限大の極限における漸近的な挙動を解析することで、グラフ CNN の表現能力とグラフの位相的情報を関連付ける方法を提案する。本講演の結果は Oono and Suzuki (2019, arXiv:1905.10947) に基づく。

$G = (V, E)$ をノード数を N の無向グラフとする。本研究では G 上のメッセージパッシング型グラフ CNN を考える。 $P \in \mathbb{R}^{N \times N}$ を対称行列、 $W_l \in \mathbb{R}^{C \times C}$ ($l \in \mathbb{N}$) とし、第 l 層目の変換 $f_l: \mathbb{R}^{N \times C} \rightarrow \mathbb{R}^{N \times C}$ を $f_l(X) := \sigma(PXW_l)$ で定義する。ここで、 $\sigma: \mathbb{R}^{N \times C} \rightarrow \mathbb{R}^{N \times C}$ は要素ごとに ReLU 関数 $\sigma(x) := \max(x, 0)$ を適用する関数である。 G の各ノード上に C 次元ベクトルを配置した信号 $X \in \mathbb{R}^{N \times C}$ を入力とし、 L 層からなるグラフ CNN $f: \mathbb{R}^{N \times C} \rightarrow \mathbb{R}^{N \times C}$ を $f = f_L \circ \dots \circ f_1$ で定義する。

P の定義はグラフ CNN によって異なるが、グラフ G の隣接構造から決まる行列で、各ノード上の信号を集約する操作に対応する。 P の最大固有値の固有空間を U 、重複度を M とし、 U の直交基底として各元の全成分が非負のものが取れると仮定する。一般的に利用されているグラフ CNN の一種である GCN (Kipf and Welling, 2017) は、 $P := I_N - \tilde{\Delta}$ とすることでこの仮定を満たす。ここで、 $\tilde{\Delta}$ は V の各ノードに自己ループを 1 本足したグラフ \tilde{G} の正規化ラプラシアン $\tilde{\Delta} := I_N - \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}}$ である (\tilde{A}, \tilde{D} はそれぞれ \tilde{G} の隣接行列と次数行列である)。一方、 W_l は各ノード上の信号に共通の線形変換を施す重み行列に対応する。本来 $(W_l)_l$ は学習により決定するパラメータであるが、本研究では $(W_l)_l$ は既に何らかの方法で与えられているものとする。

本研究 1 つ目の主定理は層数無限大でのグラフ CNN の漸近的挙動に関するものである。 P の固有値を $\lambda_1 \leq \dots \leq \lambda_{N-M} < \lambda_{N-M+1} = \dots = \lambda_N$ とし、 $\lambda := \sup_{n \in [N-M]} |\lambda_n|$ とする。 l 層目の重み行列 W_l の最大特異値を s_l とし、 $s := \sup_l s_l$ とする。さらに、 $\mathcal{M} := U \otimes \mathbb{R}^C$ とし、 $X \in \mathbb{R}^{N \times C}$ と \mathcal{M} との距離を $d_{\mathcal{M}}(X) := \inf\{\|X - Y\|_{\text{F}} \mid Y \in \mathcal{M}\}$ で定義する ($\|\cdot\|_{\text{F}}$ はフロベニウスノルムである)。

Theorem 1 $d_{\mathcal{M}}(X_l) = O((s\lambda)^l)$ 、特に $s\lambda < 1$ ならば、 X_l は $l \rightarrow \infty$ で \mathcal{M} に指数的に近づく。

もしグラフ CNN が非線形変換 σ を持たなければ、同様の主張は容易に示される。本定理の要諦は中間層で非線形変換を行ったとしても、同様のレートが達成されることである。 $X \in \mathcal{M}$ において、ノード $i, j \in V$ の次数が等しく、かつ同じ連結要素に含まれるならば、 $X_i = X_j$ となることに注意されたい。つまり、本定理は「グラフ CNN の重みの最大特異値がグラフ G のスペクトルから定まるしきい値 $\mu := \lambda^{-1}$ よりも小さければ、グラフ CNN の出力は層数無限大の極限において連結成分と次数以外にノードを区別する情報を持たない」と解釈できる。

本研究の 2 つ目の主定理は前定理のしきい値 μ の大きさに関するものである。Erdős-Rényi グラフ $G_{N,p}$ 、すなわち N 個のノードを持ち、相異なるノード間には独立に確率 p で枝が張られるランダムグラフ上で定義されたグラフ CNN を考える。スペクトルの漸近挙動を解析することで、Erdős-Rényi グラフが十分大きくかつ密ならば、 μ は $N \rightarrow \infty$ で無限大に発散する量である事がわかる。直感的にはこの定理は「密なグラフではグラフ CNN の集約操作によりノード上の信号が混ざりやすいため、信号が均質になりやすい」と解釈できる。

Theorem 2 Erdős-Rényi グラフ $G_{N,p}$ 上のグラフ CNN を考える。 $p = \Omega(\log N/N)$ と仮定する。ある定数 $C > 0$ が存在し、十分大きい任意の N と任意の $\varepsilon > 0$ に対し、少なくとも $1 - \varepsilon$ の確率で $\mu > C \sqrt{\frac{Np-p+1}{\log(N/\varepsilon)}}$ である。