非加法的集合関数に基づくニューラルネットワークの設計とモデル解釈技術

本田あおい　(九州工業大学情報工学部)*

多変数データ解析で用いられる重回帰モデルは, 広範な現象を記述できる古典的かつ強力な線形モデルである. このモデルの非線形への拡張である包除積分モデルは, 表現力が大きく向上するだけでなく重回帰分析と同等の手法でデータ分析が可能となる扱いやすいモデルである. しかし入力次元が大きすぎる場合や変数間に多重共線性が発生した場合などパラメータが決定できなくなるという問題がある. 解決策の一つが, ニューラルネットワークで用いられる誤差逆伝播法を用いてパラメータの近似解を求めることである. つまり, 包除積分の数理モデルをニューラルネットワークと捉える, あるいはニューラルネットワークの一部として組み込んで, 通常のニューラルネットワークの学習手法を用いるというものである. 必ず何らかの近似解が求まり, さらにデータの前処理部分もネットワークに組み込むことで前処理を自動化できるというニューラルネットワークの利点も活用することができる. 逆に, 包除積分モデルを用いたネットワークを解析することで, ニューラルネットワークのブラックボックス問題を解決する一つの手段にもなりうる. ニューラルネットワーク (深層学習ネットワーク) は巨大な合成関数であるので性質のわかっている関数の組み合わせ, あるいは部分的な埋め込みでネットワークの構築ができれば解釈の可能性が広がるのではないだろうか. 解釈性を持つネットワークであれば, パラメータの初期値を適切に定めることができ学習速度を改善することができるなど, より効率的な学習手法を開発することが期待できる.

定義 1 (メビウス型包除積分モデル Cf.[1, 2])　入力データ $\boldsymbol{x}$ は $X = \{1, \ldots, n, \ldots, N\}$ の N 個の数値, $\boldsymbol{x} = (x_1, \ldots, x_n, \ldots, x_N) = \mathbb{R}^N$ とする. メビウス型包除積分モデルは次のように定義される統合関数である:

$$y(\boldsymbol{x}; \boldsymbol{w}, \otimes) \quad := \quad \sum_{A \in \mathcal{P}(X)} w_A \left(\bigotimes_{n \in A} x'_n \right) := \sum_{A \in \mathcal{P}(X)} w_A \left(\bigotimes_{n \in A} h(a_n x_n + b_n) \right).$$

$\boldsymbol{w} = \{w_A\}_{A \in \mathcal{P}(X)}$ がパラメータで, これが包除積分に用いる測度のメビウス変換に相当している. 測度には加法性を仮定しない. ここで, $\mathcal{P}(X)$ は X の部分集合全体, 関数 h はシグモイド関数やランプ関数 (最大値 1 で打ち切る) などの単調な $[0, 1]$ 区間に値をとる活性化関数である. $\otimes$ は 2 項演算から自然に定義される多項演算で $\bigotimes_{i \in \emptyset} x_i := 1$ とする. 演算 $\otimes$ は何でもよいが, 本来の包除積分は t-ノルム, つまり掛け算型の演算を使うと, 統合関数は単調性が保存されるなどの積分型のよい性質を持つことがわかっている. 通常この性質を包除積分ネットワークにも反映させたいので, 演算は t-ノルムを基本とする.

例 2 (t-ノルムの例) $\otimes$ には t-ノルムのような掛け算型の演算を使用すると, 包除積分がよい性質を持つことが分かっている. 以下が代表的な t-ノルムである.

$$\text{(i) } x \wedge y := \min(x, y) \quad \text{(ii) } x \otimes^{\mathbf{P}} y := xy \quad \text{(iii) } x \otimes_\lambda^{\mathbf{DP}} y := \frac{xy}{\max(x, y, \lambda)}, \ 0 \le \lambda \le 1.$$

* e-mail: aoi@ai.kyutech.ac.jp

(iii) のようなパラメトリック t-ノルムを含め, 色々なものが提案されており, これらの中から問題に合わせて選ぶことができる.

包除積分に用いる非加法的測度は単調な測度である. これをネットワークに反映させるためには非加法的測度に相当するパラメータ $\boldsymbol{w}$ の個々の取る値を他のパラメータに合わせて単調性の制限条件を設ける必要がある. 今回の実験には説明変数と目的変数の間に単調性が仮定されるようなデータセットを用いるが, 非加法的測度の単調性の制限はつけていない.

例 3 $X = \{1, 2, 3\}$ のとき, メビウス型包除積分モデルを書き下すと

$$\begin{aligned} y(\boldsymbol{x}; \boldsymbol{w}, \otimes) &= w_{\{1\}}x_1 + w_{\{2\}}x_2 + w_{\{3\}}x_3 + w_{\{1,2\}}(x_1 \otimes x_2) + w_{\{1,3\}}(x_1 \otimes x_3) \\ &\quad + w_{\{2,3\}}(x_2 \otimes x_3) + w_{\{1,2,3\}}(x_1 \otimes x_2 \otimes x_3) + w_{\emptyset} \end{aligned}$$

入力の次元が増えると, このモデルの項数は指数関数的に増大する. ニューラルネットワークでは w_A がエッジに, $\bigotimes_{n \in A} x_n$ がユニットに相当する. ニューラルネットワークでは過学習を防ぐためのモデルを単純化するために, Dropout (Hinton 2014) によるユニッドの無効化や, ランダムまたは正則化関数を用いたエッジのスパース化が提案されている. 包除積分ネットワークでもこれらの手法を使うことは可能であるが, さらに積分に用いる測度の非加法性を制限したモデルの自由度を制限することができる.

定義 4 (k 次加法的測度 (Grabisch1998, Cf. [3])) $(X, \mathcal{P}(X))$ 上の集合関数 μ に対して, 集合関数 $m : \mathcal{P}(X) \to \mathbb{R}$ とある自然数 k が存在し,

$$\mu(A) = \sum_{B \subset A, |B| \leq k} m(A)$$

が成り立つとき μ を k 次加法的という.

例えば例 3 の場合で 2 次加法的に制限すると, 第 7 項が削除される. このように制限により項数を自由に調整して単純化することができる. これは包除積分ニューラルネットワークモデルでは $|A| > k$ に対して $w_A = 0$ としてエッジを無効化することに相当し, ランダムにユニット数を半分程度に半減させる Hinton の Dropout に類似した調整手法とも言える. ちなみに, k 次加法を用いる場合は $\lceil \frac{N}{2} \rceil$ 次加法程度にするとユニット数はほぼ半減する.

本講演では, 1 つのメビウス型包除積分を表現した単純メビウス型包除積分ニューラルネットワークを用いたデータ解析手法とネットワークの設計を紹介し, 回帰問題のデータセット (糖尿病のデータセット) を用いて, 有効性の検証実験について通常のディープラーニングネットワークを含む主要なデータ解析手法との比較の報告, 及び提案ネッワークの解釈手法を提案する.

参考文献

[1] A. Honda, Y. Okazaki, Theory of inclusion-exclusion integral, *Information Science*, **376**, pp. 136-147, 2017.

[2] 本田あおい, 包除積分数理モデルを用いた多変量データ解析, 知能と情報 (日本知能情報ファジィ学会誌) 解説論文, **30**, (4), pp. 183-192, 2018.

[3] A. Honda, R, Fukuda, Y. Okazaki, Non-discrete k-order additivity of a set function and distorted measure, preprint.