ISSN 2186-5647

P-31

Kolmogorov–Arnold Network における計算効率向上のためのゲート機構の導入

日大生産工　○赤羽　真之亜　　日大生産工　山内　ゆかり

1.　まえがき

多層パーセプトロン(MLP)はディープラーニングモデルの基盤となる構成要素である。MLP の重要は言い尽くせないほどあり、普遍的近似定理によって保障されるその表現力のおかげで、非線形関数の近似において機械学習のデフォルトモデルとなっている。しかし MLP には問題点がある。MLP はノードに固定の活性化関数を持っており、これにより全体の非線形性が限定され、非常に複雑な関数の表現力に限界が生じる。また MLP は線形重み行列と活性化関数の繰り返しによって非線形関数を近似するが、各重み行列は線形変換であるため、MLP が本質的に非線形な関数を表現するためには多くの層やノードが必要となる。つまり、複雑な関数を表現するためにはネットワークのサイズが大きくなり、パラメータ数が膨大となる。

そこで提案されたのが Kolmogorov–Arnold Networks(KAN)である。KAN は少ないパラメータで高い表現力ができる特性を持っている。さらに KAN のアーキテクチャは柔軟であり、さまざまなタスクやデータセットに適応しやすい適応性も兼ね備えている。これらの特性により、特定のタスクにおいて MLP よりも優れた性能を示すことができる。しかし KAN には制御点やノット数を増やすと、モデルがより複雑になり、不要な情報が多いという問題がある。

本研究では，この問題を解決するために gKAN(ゲート付きKAN)を提案する。MNIST データセットを用いた計算実験により通常の KANと提案手法であるgKANとの精度、処理時間を比較し報告する。

2.　従来研究

2. 1　Kolmogorov–Arnold Network（KAN）

KANはコルモゴロフ・アーノルド表現定理を基に任意の多変数関数を少数の一変数関数の組み合わせで表現する数学的なフレームワークである。

2. 1. 1　コルモゴロフの定理

コルモゴロフの定理は任意の連続関数fが、いくつかの一変数関数の和として表現できることを示しており、具体的には$f:[0,1]^n \to \mathrm{R}$という関数が以下の形で表現できる。

$$f(x)=\sum_{i=1}^{2n+1}\phi_i\left(\sum_{j=1}^{n}\phi_{ij}(x_j)\right) \tag{1}$$

ここで、ϕ_iは外部変数を表しておりϕ_{ij}は内部変数を表している。これにより、任意の複雑な多変数関数を小数の一変数関数の組み合わせとして表現できることで、次元削減を効果的に行える。

2. 1. 2　B–Spline曲線

B–spline曲線とは与えられた複数の制御点とノットベクトルから定義される滑らかな曲線である。以下がB–splineの多項式である。

$$spline(x)=\sum_{i}^{n}c_iB_{i,k}(x) \tag{2}$$

ここでc_iは制御点に対応する係数、$B_{i,k}(x)$はB–spline基底関数、kはスプラインの次数である。

基底関数$B_{i,k}(x)$は以下のように定義されている。最初に次数k=0の基底関数は単純な階段関数として以下のように定義されている。

$$\begin{cases} 1 & if\, t_i \le x < t_{i+1} \\ 0 & otherwise \end{cases} \tag{3}$$

次に次数k>0の基底関数は次の再帰関数を用いる。

$$\begin{aligned} &B_{i,k}(x) \\ &=\frac{x-t_i}{t_{i+k}-t_i}B_{i,k-1}(x) \\ &+\frac{t_{i+k+1}-x}{t_{i+k+1}-t_{i+1}}B_{i+1,k-1}(x) \end{aligned} \tag{4}$$

ここでt_iはノットベクトル（制御点に対応する点）の値である。

Introducing MobileNet into Kolmogorov-Arnold Network

Manoa AKABA and Yukari YAMAUCHI

2. 2　gMLP

gMLP はトークン間の空間情報を取得するために様々な工夫が行われている。そのために gMLP-block を構築し、それを何層か積み重ねることで適切な特徴量を獲得できる。この gMLP-block における工夫された点が Spatial Gating Unit(SGU)である。Figure1 に gMLP の構造を示す。

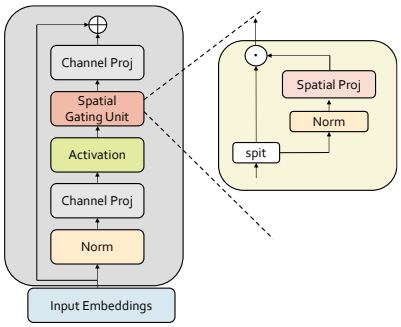

Figure 1 SGU を使用した gMLP の構造

gMLP-block は線形射影と活性関数を組み合わせて出力した値に対して SGU を用いることで、トーク間の空間情報取得している。その後、再び線形射影を行うことで必要な特徴量を獲得している。

3.　提案手法

KAN はもともと少ないパラメータで高い表現力を持つが、制御点やノット数が増加すると不要な情報が多いという問題があった。そこで今回 KAN にゲート機構を組み込むことで、情報の流れを制御することを目指す。gKAN では、ゲート機構を導入することで出力を動的に制御し、過学習を防ぎつつ計算の安定性を向上させる。Figure2 にこの手法の構造を示す。

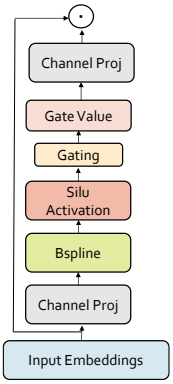

Figure 2 提案手法の構造

最初のチャネル投影では入力データに対して線形変換を行い、特徴を次の層に渡す。続いて、Bspline 基底関数を使用して入力特徴を Bspline 関数で変換し平滑な補間を行う。これは gKAN の特徴的な部分である。次に、SiLu 活性化関数が Bspline 関数の出力に適用され、非線形性を導入する。その後、ゲーティングユニットが出力を調整するためのゲート機構を提供し、別の重みや Bspline を通じて計算された値（ゲート値）を用いてメインの出力を修正する。最後にメインの出力に対してゲート値を掛け合わせることで、最終的な出力を得る。

4.　実験および検討

データセットはMNISTを使用し、従来手法と提案手法の精度、計算速度について記録を取る。実験では学習率0.0005、制御点2、ノット数6、トレニーングデータ2000、テストデータ500とする。

MNISTデータでの従来手法と提案手法の実験結果を表1に示す。

表 1. 実験結果

	訓練精度	テスト精度	実行時間
従来手法	0. 702	0. 624	812
提案手法	0. 785	0. 754	800

従来手法よりも提案手法の方が訓練精度、テスト精度、実行時間の全てで上回っている。このことから提案手法は有効であると言える。

5.　まとめ

本研究では、KANにゲート機構を加える提案を行った。KANの非線形カーネル関数で得られた出力にシグモイド関数を使ったゲート機構を適用することにより動的な出力制御が行われ、カーネル関数による複雑な非線形変換に対して柔軟に対応することにより精度の向上と実行時間の短縮を示した。今後、KANの制御点とノット数を動的に変化させるグリッド拡張やスパース化、プルーニングを行い性能向上や学習の安定性を目指す。

参考文献

1）Ziming Liu, Yixuan Wang, Sachin Vaidya, Fabian　Ruehle, James Halverson,　Marin Soljačić, ThomasY. Hou, Max　Tegmark “KAN：Kolmogorov-Arnold Networks”, arXiv：2404. 19756v1 30 Apr 2024

2）Hanxiao Liu,　Zihang Dai,　David R.　So, Quoc V.　Le　“Pay Attention to MLPs”, arXiv：2105. 08050v2 17 May 2021