

Kolmogorov–Arnold Network のマーケティング 解析への応用可能性の検討

—従来の深層学習手法との理論的比較と実データによる購買予測への応用—

新美 潤一郎[†]

[†]名城大学 経営学部

E-mail: jniimi@meijo-u.ac.jp

概要

近年、深層学習をはじめとした機械学習手法は急速な進化を続けている。例としてメディアデータの生成が可能な拡散モデルや、自然なコミュニケーションが可能な大規模言語モデルが登場しており、幅広い分野において大きな影響を及ぼしている。また、従来の深層学習とは異なるアーキテクチャーをもつモデルとして、Kolmogorov–Arnold Network (KAN) が注目され、それをベースとしたモデルが立て続けに提案されている。しかしながら、社会科学における応用は必ずしも進んでいるとはいえ、その応用可能性は不明瞭である。そこで本研究では、まずここまでの深層学習の発展について網羅的に俯瞰した上で、KAN を使用した将来購買の予測を行うとともに、モデル探索を通じて各パラメーターの効果について検証する。また、KAN の有用性の一つである活性化関数の可視化を通じて、マーケティング分析への応用可能性を検討する。

キーワード: 機械学習, 深層学習, Kolmogorov–Arnold Network, マーケティング, 購買予測

KAN in Marketing: Application of Kolmogorov–Arnold Network to Purchase Prediction in E-Commerce

—Theoretical comparison with conventional deep learning models—

Junichiro Niimi[†]

[†]Faculty of Business Management, Meijo University

E-mail: jniimi@meijo-u.ac.jp

Abstract

In recent years, deep learning has evolved rapidly, making a significant impact across a wide range of fields. For examples, generative models can produce media data, including audio and video, while large-scale language models are capable of communicating with us in the natural way. In addition, a model with a different architecture from conventional models, referred to as Kolmogorov–Arnold Network (KAN), has attracted attention. However, its application in the social sciences has been limited and remains unclear. Therefore, in this study, we first provide a comprehensive overview of the development of deep learning up to this point, then predict future purchases using KAN, and verify the effect of each parameter through model exploration. Furthermore, by visualizing the activation function which is one of the usefulness of KAN, we examine the potential application in marketing analysis.

Keywords: Machine Learning, Deep Learning, Kolmogorov–Arnold Network, Marketing, Purchase Prediction

1 はじめに

近年、深層学習をはじめとした機械学習手法は急速な進化を続けている。例として画像や動画などのメディアデータの生成までが可能な拡散モデル (Ho et al., 2020) や、自然なコミュニケーションの可能な大規模言語モデル (Large Language Models, LLM) が登場しており、単なるアプリケーション等の創出にとどまらず、社会全体の働き方や意思決定プロセスに変革をもたらすなど、幅広い分野において大きな影響を及ぼしている。特に Transformer (Vaswani et al., 2017) やそれに基づく BERT (Devlin et al., 2019) 等から、アーキテクチャーの複雑化とパラメーターの大規模化を伴う発展が続いている。

マーケティング分野でも、企業では自社顧客に関する行動ログの収集基盤の構築が進むとともに、顧客へのパーソナライゼーション等のための分析が発展してきた。伝統的には、戦略策定への活用を目的として解釈可能性の高いパラメトリックなモデルが好まれるが、その高い予測性能から機械学習も盛んに取り入れられている。主要なタスクとしては、情報推薦の一環として、過去に収集された行動や嗜好からのパーソナライゼーションに基づくターゲティングや、製品・サービスのレコメンデーションがある (Ma and Sun, 2020)。特に、顧客の行動的／態度的なロイヤルティの高さを適切に把握することは、顧客関係管理 (Customer Relationship Management, CRM) の文脈で重要な課題であり (Jacoby and Chestnut, 1978)、それらの情報予測のために機械学習手法は活用されてきた。また、自然言語処理 (Natural Language Processing, NLP) の一環としての、オンライン上の電子口コミ (electronic word-of-mouth, eWOM) を用いたセンチメント分析等は、特に機械学習を用いることに高い有用性がある (Yadav and Vishwakarma, 2020)。また、近年では価格の最適化やサプライチェーンの管理など、より幅広い領域で用いられている (Ma and Sun, 2020; Ngai and Wu, 2022)。しかしながら、そこにおける課題はやはり解釈可能性である。つまり、深層学習をはじめとした主要な機械学習モデルはその多くが非常に複雑な構造をもつがゆえに、解析結果に対する説明性が著しく低いことや、収集されたデータのバイアスをどの程度反映しているかが不明瞭であることなどの問題が指摘されてきた (Volkmar et al., 2022)。

このように、機械学習分野の発展とそのマーケティング応用は、必ずしも互いに呼応しあって進んできたとは言いがたい。しかしながら、これまでの深層学習の多くが順伝播型ネットワーク (Feed-Forward Neural Network, FFNN) や Multi-Layer Perceptron (MLP; Rumelhart et al., 1986a) 等*1 を基本に発展してきたのに対して、それとは全く異なる構造をもつ Kolmogorov-Arnold Networks (KAN; Liu et al., 2024) が提案され、深層学習分野で大きな関心を集めている。KAN では、そのアーキテクチャーが従来の MLP と異なるだけでなく、各変数が目的変数に与える影響についても可視化することが可能であるという意味で、従来手法より高い解釈可能性を持っているとされる (Liu et al., 2024)。もちろん、KAN は提案されて間もないモデルであり、特に社会科学分野における応用はほとんどないことから、その応用可能性は未知数である。

そこで本研究では、KAN やその理論的背景について整理した上で、マーケティング解析への応用を行い、その解析例からモデルの有用性について検討する。以降の構成として、第 2 節ではここまでの深層学習の発展について俯瞰し、第 3 節で理論的背景から KAN を導入する。それらに基づいて第 4 節で提案モデルを構築し、実データを用いた解析とモデルの可視化を行う。最後に、第 5 節で解析から得られた示唆と本研究の限界をまとめる。

*1 本研究では再起型や畳み込み型を含めさまざまな構造の深層学習を整理するため、特にノード間の計算の方向が重要となる場合に FFNN という表記を用いる。

2 深層学習領域の俯瞰的整理

KAN の構造を整理するのに先立ち、これまでの深層学習の流れを確認する。そもそも深層学習 (Deep Learning; Bengio et al., 2017) は、潜在変数を階層的に重ねた深層構造と、非線形関数を用いた活性化を組み合わせたアーキテクチャーを持つ。現在の深層学習において最も基本的なモデルであると考えられる MLP*² では、特徴抽出により目的変数に関連する複雑な特徴量の学習が可能であり、従来の機械学習手法と比較して非線形関数への高いフィッティング能力を示す。主な活性化関数として sigmoid 関数や rectified linear unit (ReLU)、hyperbolic tangent (tanh) 等が用いられる。

のちの KAN の構造と対比するため、ここで MLP の順伝播計算の構造について概説する。 L 層の隠れ層と、入力層 $l = 0$ を含む層 $l = \{0, 1, 2, \dots, L\}$ に J_l 個のニューロンを持つ MLP において、第 l 層 j_l 番目 (ここで $j_l = \{1, 2, \dots, J_l\}$) のニューロンの状態 h_{l,j_l} から第 $l+1$ 層 j_{l+1} 番目のニューロンの状態 $h_{l+1,j_{l+1}}$ への計算は、パラメーターとして h_{l,j_l} から $h_{l+1,j_{l+1}}$ へのウェイト $w_{j_l,j_{l+1}}$ と層全体のバイアス b_l 、各層 l について固定的に設定された活性化関数 $\phi_l(\cdot)$ を用いて、

$$h_{l+1,j_{l+1}} = \phi_l \left(\sum_{j_l=1}^{J_l} w_{j_l,j_{l+1}} h_{l,j_l} + b_l \right) \quad (1)$$

となる。つまり上位の隠れ層の状態計算は隠れ層 l の全ニューロン $\mathbf{h}_l = (h_{l,1}, h_{l,2}, \dots, h_{l,J_l})$ のアフィン写像と非線形活性化の組み合わせによって行われる。行列表現とするならば内積を用いて

$$\mathbf{h}_{l+1} = \phi_l(\mathbf{W}_l \mathbf{h}_l + \mathbf{b}_l) \quad (2)$$

である。さらに MLP 全体では、入力層を $l = 0$ とすれば $\mathbf{h}_0 = \mathbf{x}$ より、

$$MLP(\mathbf{x}) = \phi_L(\mathbf{W}_L \cdots \phi_2(\mathbf{W}_2 \phi_1(\mathbf{W}_1 \phi_0(\mathbf{W}_0 \mathbf{x} + \mathbf{b}_0) + \mathbf{b}_1) + \mathbf{b}_2) \cdots + \mathbf{b}_L) \quad (3)$$

となる。

マーケティング応用では、スマートフォンの非契約型かつ Free-to-Play (F2P) なゲームアプリにおける顧客生涯価値 (Customer Lifetime Value, CLV) としてのアプリ内購買額を MLP により予測した研究 (Chen et al., 2018) や、購買予測において MLP と回帰分析それぞれでその精度や特性を比較し MLP の比較的単純な特徴抽出手法の限界を指摘した研究 (新美・星野, 2017b) がある。特定の分布を仮定しないフィッティングであることから、時間や金額などマーケティング分野の中でも比較的大きなスケールの変数を対象とした解析に用いられることが多い。

その後は、画像等の 2 次元データを想定した Convolutional Neural Network (CNN; LeCun et al., 1995) や、時系列データへの適用を目的として再帰的な構造をとる Recurrent Neural Network (RNN; Elman, 1990; Jordan, 1997) などデータの形状に合わせたアーキテクチャーの設計に、その発展系としての Long Short-Term Memory (LSTM; Hochreiter and Schmidhuber, 1997) や Gated Recurrent Unit (GRU; Cho et al., 2014) 等が続いた。画像処理分野は CNN から深層化した VGG-16 (Simonyan and Zisserman,

*² もちろん MLP の確立には、backpropagation (Rumelhart et al., 1986b) から、Restricted Boltzmann Machine (RBM; Smolensky et al., 1986; Hinton and Salakhutdinov, 2006; Larochelle and Bengio, 2008)、さらには Deep Belief Network (DBN; Hinton et al., 2006)、あるいはその学習のための greedy layer-wise training (Bengio et al., 2006) までの貢献が欠かせない。

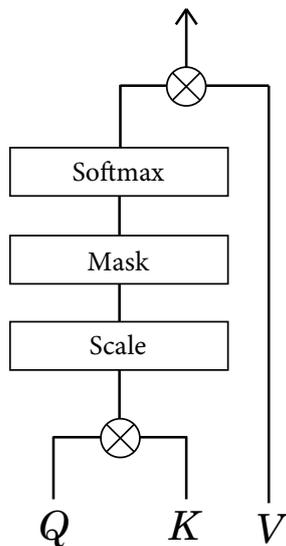


図1 Attention Mechanism

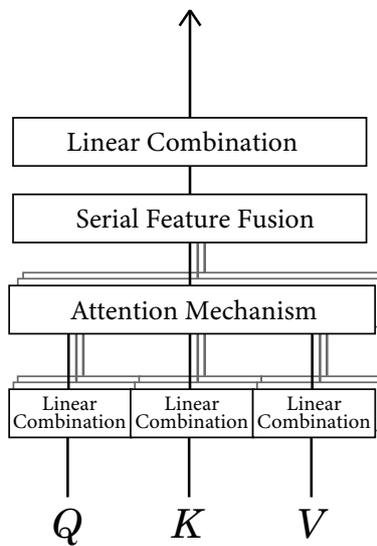


図2 Multihead Attention

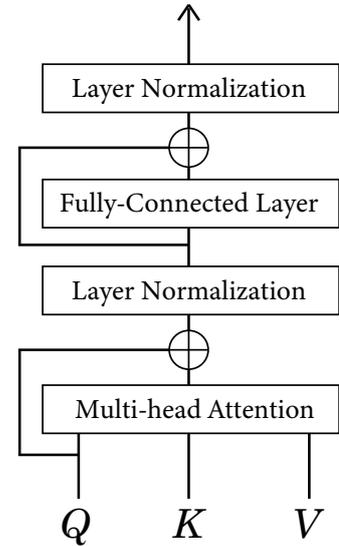


図3 Transformer Encoder

2014) や Inception モジュール (Szegedy et al., 2015)、ノード間の接続の工夫としての残差接続 (Residual Connection; He et al., 2016) 等により発展してきた。

マーケティング分野において解析に用いる行動ログはパネルデータ (多エンティティ多変量時系列データ) として得られることが多いため、RNN やその発展モデルを用いた応用研究が盛んである。例として、EC サイトで収集された Clickstream data を用いて、あるセッションが購買・ショッピングカートの放置・閲覧のみのどの状態で終了するかを予測した研究 (Toth et al., 2017) や、金融サービスの利用状況を月次レベルの RFM に要約した時系列データを用いて、LSTM により離脱を予測した研究 (Mena et al., 2019)、複数の RNN 系手法による Clickstream data からの購買予測にあたり MLP や Random Forest 等とその精度を比較した研究 (Koehn et al., 2020)、さらに時系列データの解析における単位時間 (時系列的な粒度) が予測精度に及ぼす影響を検証した研究 (新美, 2021) 等がある。また CNN を用いたものには、EC サイトからデータを収集し、地域・カテゴリ・ブランド・商品について作成した時系列データを 4 チャンネルの 2 次元データとみなして販売量を予測した研究 (Pan and Zhou, 2020) や、従来のな構造化データに加えて、金融サービスプロバイダーにおける顧客とアドバイザーの間で交わされたメッセージ文書のテキスト情報を投入して、サービスからの顧客の離脱を予測した研究 (De Caigny et al., 2020) 等がある。

それらデータの形状に合わせた構造の設計がひと段落すると、そこからは特徴抽出の方法によるブレイクスルーが続いた。例として、attention mechanism (図 1; Bahdanau et al., 2015) では、入力から生成される Query、Key、Value に基づき、特徴量の重要度に加重をかけることにより、情報を選択的に用いることが可能である。特に、複数の attention 機構を並列的に組み込むことにより (図 2) 構築された Transformer (図 3; Vaswani et al., 2017) は、入出力の間の大域的な依存関係を見出すことが可能であるという点で、それまでの手法と比較しても高い予測能力を示し、さまざまな分野で Transformer をベースにしたモデル提案や応用が行われてきた。^{*3}

^{*3} 図 1-3 はいずれも Vaswani et al. (2017)、新美 (2024) を参考に著者らが作成。

また、深層学習分野の重要な手法の一つにマルチモーダル深層学習 (Multimodal deep learning; Ngiam et al., 2011; Srivastava and Salakhutdinov, 2012) がある。マルチモーダル学習では、マルチソースデータや異なる種類のモダリティを組み合わせて予測や分類等のタスクを行う。深層学習による一般的なマルチモーダル学習は、各データセットに固有の隠れ層を用いた処理を行った上で、より上位の隠れ層でそれらの特徴量を serial feature fusion により横方向に結合し joint representation とする intermediate fusion を用いて行われる (Ramachandram and Taylor, 2017)。異なるモダリティからの情報の組み合わせや関係性に基づく特徴量の作成で、より頑健なモデル構築が可能である (Ngiam et al., 2011)。

マーケティング領域では統計的データ融合 (Statistical data fusion; Kamakura and Wedel, 1997) の文脈で似たような試みが複数存在する*4。例として、Deep Boltzmann Machine (DBM; Salakhutdinov and Hinton, 2009) を欠損データに対応するよう拡張した発展モデルにより購買履歴／調査回答のマルチソースデータから情報予測を行った研究 (新美・星野, 2017a) や、企業の外部／内部の情報をマルチソースデータとして構造化／非構造化データのそれぞれを最適なアーキテクチャーにより処理しながら LSTM により時系列的な情報予測を行う研究 (Pan et al., 2020) がある。また、明示的なマルチモーダル学習として、動的な特徴量としてのオンライン上の閲覧履歴と、静的な特徴量としての顧客のプロフィール情報のマルチモーダル学習により購買を予測した研究 (Kim et al., 2023) がある。

しかしながら、それら単純な特徴融合においては、各モダリティから出力された特徴量が全て同じウェイトで評価されることから、モデル規模の増大に対して必ずしも効率的な解析が行えないという問題が指摘されている (Niimi, 2024b)。先述の Transformer はマルチモーダル学習にも多大な貢献をしており、Attention や Transformer を用いることにより、一方のデータセットの内容に基づき他方のデータセットから情報を選択的に抽出する cross attention 等を用いたモデルも数多く提案されている (e.g., Choi et al., 2018; Yoon et al., 2019; Zadeh et al., 2018; Niimi, 2024b)。マーケティング分野においては、月次集計のクレジットカードの利用履歴を Bidirectional LSTM (Bi-LSTM; Graves and Schmidhuber, 2005) に入力した上で、Attention 機構により顧客のデモグラフィック情報とマルチモーダル学習することにより顧客のスコアリングを行った研究 (Ala'raj et al., 2021) や、スマートフォンでの EC アプリの利用にあたり、閲覧／操作履歴それぞれの階層的な attention 構造により購買意図をリアルタイムに予測した研究 (Guo et al., 2019)、さらに日次単位で集計したアプリ利用ログとユーザーの調査回答データを Source-Target Attention Transformer と特徴融合によりマルチモーダル学習した研究 (新美, 2024) がある。このように、マーケティング領域のマルチモーダル学習では、明確に画像等のメディアデータと組み合わせるというよりは、同じ tabular data であっても異なる状況や手法により収集され統合されたデータを融合して予測に用いることが多い。明確に異なるモダリティの同時学習という点では、BERT で分散表現化したオンラインプラットフォーム上のユーザーレビューの文書とユーザープロフィールを Cross Attention Transformer で融合し、顧客のレストラン評価を予測した研究 (Niimi, 2024b) や、SNS に投稿されたテキストと画像の Source-Target Attention による同時学習から、災害時に有用な情報かどうかを分類した研究 (Khattar and Quadri, 2022) 等がある。

また、Transformer のもう一つの大きな貢献に、言語モデルの発展がある。これまで自然言語処理 (Natural language processing, NLP) 分野では、テキストをいかにベクトル空間に写像し解析に投入するかが課題であり、それに関してさまざまな手法が提案されてきた。代表的なものに Bag-of-Words (Joachims, 1998) や term frequency-inverse document frequency (TF-IDF; Sparck Jones, 1972)、word2vec (Mikolov et al.,

*4 ここで「統計的」と表記されるのは、特に情報工学分野において深層学習等を用いた multisensor data fusion という類似した領域が存在するためである。

2013) / Doc2vec (Le and Mikolov, 2014) 等がある。しかしながら、これら比較的シンプルな手法においては、文章中での他の単語との関係性としての文脈 (context) を考慮できないことにより、ある単語には常に一意の値が付与されるという意味での多義性の問題が指摘されてきた (Peters et al., 2018)。そのようななか、Transformer のエンコーダー部分のみを組み込んだ Bidirectional Encoder Representations from Transformers (BERT; Devlin et al., 2019) やその発展手法 (Liu et al., 2019; Sanh et al., 2019)、あるいはデコーダー部分のみを組み込んだ Generative Pre-trained Transformer (GPT; Brown, 2020) では、単語の意味を文脈に応じて柔軟に変化させる deep-contextualized word representation (Peters et al., 2018) を獲得できるようになった。BERT も GPT も基本的なタスクには高い精度を発揮するが、なかでも BERT は双方向性の文脈理解に基づく文書の分散表現化に、一方の GPT は単語予測に基づく自然な文書生成に強みをもつ。マーケティング関係の応用研究としては、特にアノテーションやセンチメント分析 (Van Attevelde et al., 2021) において、LLM にプロンプト (指示文) を与えて分類や回帰を行う手法が模索されている (Krugmann and Hartmann, 2024)。例として、LLM を用いたチャットサービス ChatGPT (<https://chatgpt.com>) 上での提供モデル gpt-3.5-turbo やファインチューニング済みの BERT 等を用いて、センチメント分析を含む複数の解析を通じてその有用性を検証した研究 (Wang et al., 2023) や、zero-shot の言語モデルでも Amazon Mechanical Turk (MTurk) のワーカーを平均 25% 上回る精度でタスクを実行可能であることを示した研究 (Gilardi et al., 2023)、さらにはオンラインプラットフォーム上に投稿されたレストランに関するレビュー文書の観点別センチメント分析にあたり、パラメーター数が 80 億程度の LLM で推論を複数回繰り返して結果の多数決を取ることで、700 億パラメーターのモデルの単一の推論結果よりも高精度に予測を行う頑健なモデル構築が可能であることを示した研究 (Niimi, 2024a) などがある。これらの多くにおいて、LLM は伝統的な NLP 手法である VADER (Hutto and Gilbert, 2014) 等を上回る精度を示している。また、近年では単純なコミュニケーションや回帰・分類タスクにとどまらず、テキストデータからの Information Retrieval を目的とした検索拡張生成 (Retrieval-Augmented Generation, RAG; Lewis et al., 2020) 等の提案も続いている。

このように、深層学習は 1980 年代の理論的基盤の構築と 2000 年代の計算資源の発達により、20 年ほどで爆発的な進歩を遂げた。先にも書いたとおり、基本的にはそのほとんどが MLP の構造を基本として発展的に提案されてきたものである。

3 Kolmogorov–Arnold Network の導入

3.1 Kolmogorov–Arnold Representaiton Theorem

KAN の導入にあたり、KAN が理論的に依拠するとされるコルモゴロフ・アーノルド表現定理 (Kolmogorov–Arnold Representaiton Theorem, KART; Kolmogorov, 1956; Arnol'd, 1957)^{*5} について述べておく。この定理は、 J_0 個の入力 $\mathbf{x} = (x_1, x_2, \dots, x_{J_0})$ をとる滑らかな多変数連続関数 $f(\cdot)$ が、それらの各単変数を入力にとる複数のより単純な連続関数の合成関数で表現可能であることを示したものである。具体的には、内部関数 $\phi_{q,p}$ と、外部関数 Φ_q の組み合わせであり、ここで $p = \{1, 2, \dots, J_0\}$ かつ $q = \{1, 2, \dots, 2J_0 + 1\}$ 、つまり元の非線形関数に入力される変数の数 J_0 に依存して (p, q) の上限が決定される^{*6}。複数の発展形はある

^{*5} 本研究においても KART の概要には触れるが、より詳しい解説は近年の KAN 発展モデル内のレビューを参照されたい (e.g., Schmidt-Hieber, 2021)。

^{*6} 当然ながら、ある関数を $q = \{1, 2, \dots, 2J_0 + 1\}$ で表現可能であることは、直ちに関数 Φ_q を $q = 2J_0 + 1$ まで使わなければならないことを意味しない。これは表現すべき関数の複雑さに依存する。

ものの (e.g., Lorentz, 1962; Ostrand, 1965; Sprecher, 1965)、本研究では最も基礎的な形式についてその構造を整理する。まず定理で示される関数 $f(\cdot)$ の構造は、*7

$$f(\mathbf{x}) = \sum_{q=1}^{2J_0+1} \Phi_q \left(\sum_{p=1}^{J_0} \phi_{q,p}(x_p) \right) \quad (4)$$

となっている。ここで $\phi_{q,p} : [0, 1] \rightarrow \mathbb{R}$ であり、すなわち関数 $\phi_{q,p}(x_p)$ は入力 $x_p \in [0, 1]$ を実数空間 \mathbb{R} に写像する。次に、ある q における $\phi_q = (\phi_{q,1}, \phi_{q,2}, \dots, \phi_{q,p})$ の総和をとる関数 $\Phi_q(\cdot)$ を用いる。ここで $\Phi_q : \mathbb{R} \rightarrow \mathbb{R}$ であり、実数空間上の任意の値を別の実数値に写像する。そして、それら $\Phi = (\Phi_1, \Phi_2, \dots, \Phi_q)$ の和をもって元の関数 $f(\cdot)$ を表現可能とする。つまり、本定理は連続多変数関数を複数の単変数関数の 2 段階の合成により記述可能であることを示している。Schmidt-Hieber (2021) でも指摘されているとおり、この計算過程は各変数からの入力に関数を適用しつつ合計していくという意味で、ニューラルネットワークに類似した構造として解釈することも可能である。もちろん、活性化と合計の順序が入れ替わっていることから厳密には MLP と異なるが、少なくとも、その構造からは本定理における近似表現自体が深層学習のアーキテクチャーと部分的には類似していることが理解でき、応用可能性の高さが示唆される。

とはいえ、このように非線形な連続関数に対する高い表現力を担保可能な KART が機械学習分野において長らく注目されてこなかった理由の一つに、任意の連続関数への応用可能性はあくまでも限定的だと考えられてきたという点がある。たとえば、KART では扱対象としての関数の滑らかさ (smoothness) が前提とされており (Akashi, 2001; Poggio et al., 2020)、かつ内部関数 $\phi(\cdot)$ の合成により外部関数 $\Phi(\cdot)$ への入力時点ですでに関数の滑らかさが失われてしまう可能性も指摘されている (Schmidt-Hieber, 2021)。しかしながら、本定理を深層学習的な手法に拡張した KAN を皮切りに、後述するような手法によりそれらの課題が部分的にも解決されつつある (Schmidt-Hieber, 2021; Bozorgasl and Chen, 2024)。

3.2 Kolmogorov–Arnold Network への拡張

KART の深層学習への拡張は KAN 以前にもさまざま検討されてきた (e.g., Lin and Unbehauen, 1993; Köppen, 2002; Polar and Poluektov, 2021)。それでも、KART を用いた機械学習手法が大きく注目を集めたのは、本研究でも用いる KAN (Liu et al., 2024) の提案からといえる。というのも、Liu et al. (2024) は過去の類似モデルが KART の隠れ層 2 層 (ϕ, Φ)、幅 $q = 2J_0 + 1$ 構造に固執するあまり、ニューラルネットワークの最新の知見を活用できていないことを限界として指摘している。そこで KAN では、KART の思想をベースに MLP の構造を発展させ、KART の計算構造を実証モデルとして L 層 J ニューロンに一般化することが試行されている。以降では KAN のモデル構造について Liu et al. (2024) を中心に整理する。

そもそも、KAN も FFNN と同様に順伝播型のネットワークである。ここで 2 層の隠れ層をもつ KAN を FFNN の形で解釈したアーキテクチャーを図 4 に示すと*8、単一の入力変数それぞれが非線形関数により直に活性化され、それらの線形和をとる計算の繰り返し構造であることがわかる。以降これを L 層に一般化し、ある l 層が保有するノード数を J_l とする。 l 層のあるノード h_{l,j_l} から上位の別のノード $h_{l+1,j_{l+1}}$ へのエッジで適用される活性化関数を $\phi_{l,j_l,j_{l+1}}$ とする。MLP (cf. 第 2 節) と同様に $l = \{0, 1, 2, \dots, L\}$ かつ $j_l = \{1, 2, \dots, J_l\}$ であり、隠れ層は L 層より J_0 はモデルがとる説明変数の数を表す。

*7 以降 KART の計算過程は基本的に Liu et al. (2024) を参考に作成。ただし、本研究の他の式との整合性をとるため一部の記号は変更している。

*8 Liu et al. (2024) を参考に著者らが作成。整理のため、KART でいう $q = 1$ のみ表示を濃くしている。

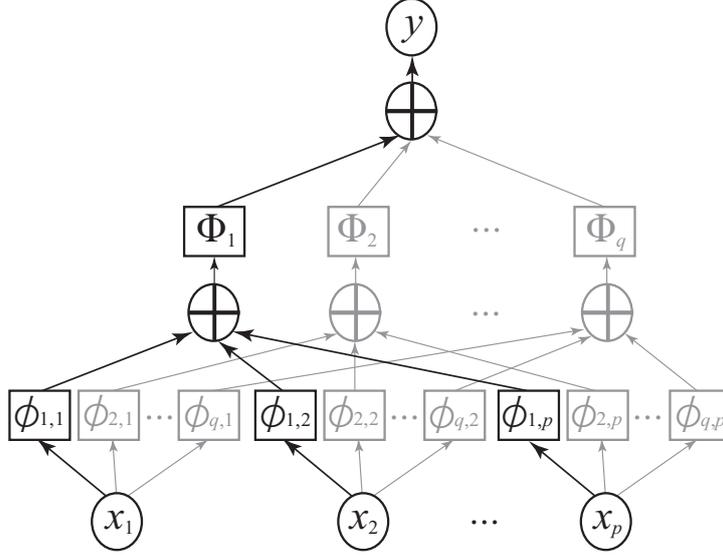


図4 KANのアーキテクチャー (浅いネットワーク)

これらの定義から、 $l+1$ 層のノード j_{l+1} への順伝播方向の逐次計算は

$$h_{l+1,j_{l+1}} = \begin{cases} \sum_{j_l=1}^{J_l} \phi_{l,j_l,j_{l+1}}(h_{l,j_l}) & (\text{if } l > 0) \\ \sum_{j_0=1}^{J_0} \phi_{0,j_0,j_1}(x_{j_0}) & (\text{otherwise}) \end{cases} \quad (5)$$

となる。つまりモデル $KAN(\cdot)$ はその階層構造より、

$$\begin{aligned} KAN(x) &= \sum_{j_L=1}^{J_L} \phi_{L,j_L,1} \left(\sum_{j_{L-1}=1}^{J_{L-1}} \phi_{L-1,j_{L-1},j_L} \left(\sum_{j_{L-2}=1}^{J_{L-2}} \cdots \sum_{j_2=1}^{J_2} \phi_{2,j_2,j_3} \left(\sum_{j_1=1}^{J_1} \phi_{1,j_1,j_2} \left(\sum_{j_0=1}^{J_0} \phi_{0,j_0,j_1}(x_{j_0}) \right) \right) \right) \right) \end{aligned} \quad (6)$$

と表現できる。これを行列表現にするには、 l 層の活性化関数行列 ϕ_l を用いて

$$\mathbf{h}_{l+1} = \phi_l \mathbf{h}_l = \begin{pmatrix} \phi_{l,1,1} & \phi_{l,1,2} & \cdots & \phi_{l,1,j_{l+1}} \\ \phi_{l,2,1} & \phi_{l,2,2} & \cdots & \phi_{l,2,j_{l+1}} \\ \vdots & \vdots & \cdots & \vdots \\ \phi_{l,j_l,1} & \phi_{l,j_l,2} & \cdots & \phi_{l,j_l,j_{l+1}} \end{pmatrix} \mathbf{h}_l \quad (7)$$

であり、したがってモデル $KAN(\cdot)$ の入出力は

$$KAN(\mathbf{x}) = (\phi_L \circ \phi_{L-1} \circ \cdots \circ \phi_1 \circ \phi_0) \mathbf{x} \quad (8)$$

となる (cf. Liu et al., 2024; Bozorgasl and Chen, 2024; Drokin, 2024)。

ここまで整理したことからもわかるように、構造が決定されてからのKANの順伝播計算は、深層学習の行列計算に慣れていればむしろ単純ともいえるものである。しかしながら、KARTからKANへの拡張にあ

たつての最大の課題はその非線形関数 $\phi(\cdot)$ の表現方法である。KART により保証されているのはあくまでも任意の滑らかな多変数連続関数を近似可能であることのみであり、実際の動的な学習構造やパラメーターの推定方法は新たに構築する必要がある。その方法はさまざま考えられるが、KAN ではスプライン関数 (Spline function; De Boor, 2003) を採用している。スプライン関数では空間を x 軸方向 K 区間に分割し、各区間内の関係性は d 次関数により表現される。ある区間 k (ここで $k = \{1, 2, \dots, K\}$) 内での x 軸上の相対位置を用いた基底スプライン関数 $B_{k,d}(\cdot)$ は

$$B_{k,d}(x) = \begin{cases} \left(\frac{x - t_k}{t_{k+1} - t_k} \right)^d & (\text{if } x \in [t_k, t_{k+1}]) \\ 0 & (\text{otherwise}) \end{cases} \quad (9)$$

となる。ここで t_k は区間 $k-1$ と区間 k の境界 (ノット) を表し、よって任意の k について常に $t_k < t_{k+1}$ である。簡単のため $d=0$ とすれば、

$$B_{k,0}(x) = \begin{cases} 1 & (\text{if } x \in [t_k, t_{k+1}]) \\ 0 & (\text{otherwise}) \end{cases} \quad (10)$$

であり、基底スプラインは区間の中で常に一定の値をとる。区間全体にわたるスプライン関数 $S_d(\cdot)$ は、ある区間 k においてのみ有効なパラメーター c_k を用いて

$$S_d(x) = \sum_{k=1}^K c_k B_{k,d}(x) \quad (11)$$

と表現可能である。隣り合う区間の関数同士はノット上で同値を取るため、スプライン関数は定義域全体にわたって連続である。

KAN の実際の活性化関数 $\phi(\cdot)$ は、式 11 に加えて、残差接続に似た関数として Sigmoid-weighted Linear Unit (SiLU; Elfving et al., 2018) とスケールパラメーター (w_b, w_s) を用いて、

$$\phi(x) = w_b \text{SiLU}(x) + w_s S_d(x) \quad (12)$$

$$\text{where } \text{SiLU}(x) = x \frac{1}{1 + e^{-x}} \quad (13)$$

としている (cf. De Boor, 2003; Liu et al., 2024)^{*9}。

3.3 MLP と KAN の比較

ここまでに KAN の根底にある KART と、モデルの基本的な設計思想について整理した。従来の MLP との差異について Liu et al. (2024) を元に整理すると、MLP ではあくまでもニューロン上に学習可能な線形パラメーター (ウェイト/バイアス) が配置されており、入力の線形加重をとった上で、事前に設定された非線形関数による活性化が行われる。一方の KAN では、ウェイトやバイアスを持たず、入力は 1 変数ごとに処理され、スプライン関数により区間ごとに学習可能な関数で表現される。そして単純な総和により上位ノードの値が決定される。それらを比較した表 1 から、従来の深層学習のアーキテクチャーとは大幅に異なる形で構築されていることがわかる。

^{*9} 式 (12) は元論文 Liu et al. (2024) Eq.(2.10) から作成しているが、当該の式は *arXiv* v4 (published in 16 Jun 2024, UTC) で修正されていることに注意されたい。

表 1 Comparison between MLP and KAN

	MLP	KAN
Expression (from h_l to h_{l+1})	$\mathbf{h}_{l+1} = \phi_l(\mathbf{W}_l \mathbf{h}_l + \mathbf{b}_l)$	$\mathbf{h}_{l+1} = \phi_l \mathbf{h}_l$
Parameters (e.g., weights, biases)	linear, trainable	none [†]
Activation functions	nonlinear, constant	nonlinear, trainable

[†] As shown in Eq.(12), KAN obviously contains the trainable scale parameters (w_b, w_s) in the activation function $\phi(\cdot)$; however, these weights are redundant as they are substantially absorbed into the SiLU and Spline functions (Liu et al., 2024).

本来の KART において近似対象となる関数は有界領域にあることが前提とされるが、KAN ではスプライン上の各区間を有界領域として区間ごとに非線形関数を推定することにより、有界領域の連続としての定義域の一般化を実現していると推察される。この構造の利点として、第一には関数表現の近似にあたっての局所的な制御が可能であることが挙げられる。また、従来の MLP では入力行列に対して線形加重を行った上でさらに活性化するが、KAN ではスプラインを用いることにより重み行列を持たず、単変数の活性化関数に直接学習させる。MLP において、活性化関数はハイパーパラメーターとして固定的に設定され、すなわち学習の過程を通じて一定である。これは最適な活性化関数の選択のために網羅的な探索が必要となることを意味している。別の利点として、深層構造でも各レイヤーにおいては常に単変数の計算となるため、解釈性が高まる点についても強調されている。

一方で、モデルの学習にあたっては、高次元データでの次元の呪い (curse of dimensionality) に陥りやすいことも指摘されている (Liu et al., 2024)。特に、スプライン関数の推定は計算負荷が高いことから、エポックあたりの計算が遅いことが指摘されている (Liu et al., 2024; Cheon, 2024)。ただし、MLP と比較してモデル規模や一般化可能性の観点で効率的である可能性から、MLP より少ないエポック数やパラメーター数で高い予測精度を達成できるという指摘もある (Lobanov et al., 2024)。実際、KAN と MLP を機械学習、NLP、数式表現など 5 つの分野で比較した研究 (Yu et al., 2024) によれば、モデル間のパラメーター数と計算性能 (FLOPs) を統制した総合評価により、性能は MLP に劣るものの、一方で総合的な学習時間や解釈性の点で優れていることが示されている。しかしながら、根本的な構造が異なる以上、パラメーター数を揃えることによる程度の意味があるかは疑わしく、むしろ計算時間を揃えた場合の予測精度の比較が重要になるのではないかと考えられる。

KAN をベースとした発展手法も 2024 年のうちに立て続けに提案されている。代表的なものに、活性化後の値をマスキングすることによる Dropout 機構の導入 (Altarabichi, 2024)、時系列分析への応用 (Genet and Inzirillo, 2024; Vaca-Rubio et al., 2024; Xu et al., 2024)、畳み込み構造の導入 (Bodner et al., 2024; Drokin, 2024)、Graph Network への拡張 (Kiamari et al., 2024; Bresson et al., 2024)、スプライン関数の wavelet 関数への置換 (Bozorgasl and Chen, 2024) やチェビシェフ多項式への置換 (Sidharth, 2024) 等がある。さらに、KAN の応用研究も少しずつ進んでいる。例として、リモートセンシング分野 (Cheon, 2024) や、スペクトラル解析 (Lobanov et al., 2024)、量子アーキテクチャ探索 (Kundu et al., 2024) に用いられている。ここに挙げたほとんどの文献は preprint 段階であることには留意が必要であるものの、一方でモデル提案から間もなくこれだけの発展モデルや応用研究が進んでいる点は強調しておきたい。

4 実データ解析

4.1 解析の概要

本研究では KAN のハイパーパラメーターの設定によりモデルの汎化性能がどのように変化するのかを網羅的な分析により明らかにする。具体的には、異なる設定の複数の KAN を用いて、ある 1 ヶ月の購買状況をもとに翌月の購買金額を予測し、その結果を比較する。解析に用いる実データには、オープンデータとして提供されている海外の大手家電量販店の ID-POS (REES46 Marketing Platform, 2020) を用いる。本データはマーケティング・プラットフォームである REES46 (rees46.com) により収集され、2020 年 4 月から 11 月までのオンラインストアでの 260 万件の商品購買が収録されている。2020 年 8 月の購買履歴から特徴量を作成し、翌 8 月の購買金額を予測する。使用する変数の詳細と要約統計量をを示した表 2 を元に、サンプルサイズ n のデータセット $D = \{(x_i, y_i)\}_{i=1}^n$ で説明変数 $x_i = (x_{i,1}, x_{i,2}, \dots, x_{i,J_0}) \in \mathbb{X}$ と目的変数 $y_i \in \mathbb{Y}$ とすれば、本研究では $n = 6000$ 、 $m = 8$ 、 $J_0 = 5$ である。標準化前の要約統計量からは、特に金額について上方向に外れ値が存在していることがわかる。

表 2 Variable Descriptions and Summary Statistics ($m = 8$)

Variable Name	Mean	Std	Min	Median	Max	Details
Target Variable						
$TotalAmount_{m+1}$	882.690	1495.986	0.440	407.360	22433.439	Total amount spent in the next month.
Explanatory Variables						
$TotalAmount_m$	1044.777	1584.217	0.900	539.300	29298.689	Total amount spent.
$nPurchases_m$	2.774	3.007	1.000	2.000	52.000	Total number of purchases.
$nCategories_m$	2.394	2.398	1.000	1.000	29.000	Total number of categories purchased.
$nBrands_m$	2.944	3.378	1.000	2.000	52.000	Total number of brands purchased.
$nProducts_m$	2.449	2.518	1.000	1.000	31.000	Total number of items purchased.

学習中の検証と学習後のテストにサンプルのうち 15% を抽出し、学習には残りの 4200 サンプルを用いる。学習にあたって early stopping (Prechelt, 1998) は適用せず 100 Epoch まで行う。ただし、最終的な予測精度は学習中に validation loss の最も低下した時点のパラメーターを用いて算出した train loss、validation loss、test loss を用い、評価には test loss (汎化誤差) を用いる。

各ハイパーパラメーターの設定が及ぼす効果を確認するため、表 3 に示す条件に基づきグリッドサーチによる探索を行う (特徴量の標準化については次節を参照)。次元数 d や区間数 K 、隠れ層の規模に加えて、主要な最適化手法として確率的勾配降下法 (Stochastic Gradient Descent, SGD; Bottou, 2012)、Adam (Kingma and Ba, 2014)、Adamax (Kingma and Ba, 2014)、AdamW (Loshchilov, 2017) を用いる。また、学習中の learning rate の調整のため、線形/指数関数型の Scheduler を導入した場合の変化についても検証する。それぞれのカテゴリカルな条件から、探索する総モデル数は 7200 件である。

さらに、提案モデルの有用性を検証するため、複数の参照モデルで提案モデルと同一の解析を行い、その精度を比較する。まず、一般的な MLP として、隠れ層にそれぞれ Rectifier Linear Unit (ReLU) $\phi^{(R)}$ と hyperbolic tangent (tanh) $\phi^{(T)}$ を採用した Feed-Forward な MLP を構築する。活性化関数の構造はそれ

表 3 Hyper-parameters for the Proposed Model

Parameters	Candidates
Constant	
Batchsize	64
# of Epochs	100 (early-stopping is not applied)
Loss Function	MSE
Evaluation	Test loss in RMSE when validation loss hits lowest
Explored	
# of Dimensions in Splines (d)	{ 1, 2, 3, 4, 5 }
# of Grids in Splines (K)	{ 10, 20, 50 }
A Size and Depth of Hidden Layer(s)	{ (3), (5), (10), (50), (5, 3), (10, 5), (20, 10, 5), (50, 20, 10) }
Optimizer	{ SGD, Adam, Adamax, AdamW }
Scheduler	{ None, Linear, Exponential }
Feature Normalization	{ None, Min–Max, Sigmoid, Percentile, Clipping Norm. }

ぞれ

$$\phi^{(R)}(x) = \max(0, x) \quad (14)$$

$$\phi^{(T)}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (15)$$

である。最適化には Adam を用い、KAN と同様に 100 Epochs の学習のうち validation loss の最も低下したモデルに対して評価を行う。またその他のモデルとして、主要な機械学習手法のうち XGBoost (Chen and Guestrin, 2016)、LightGBM (Ke et al., 2017)、Support Vector Regression (SVR; Cortes and Vapnik, 1995; Smola and Schölkopf, 2004)、k-Nearest Neighbor Regression (kNN; Cover and Hart, 1967) を用いる。

4.2 特徴量の標準化

先に述べたとおり、KAN では関数の定義域についても一般化されている。それは、先の基底スプラインに関する式 9 から理解できるように、近似対象となる関数全体の定義域がどの範囲であろうと、それらを K 区間に分割するとともに、区間 k 内における相対的な位置は $[0, 1]$ に標準化されてモデリングされる。つまり、特に入力側が外れ値を取るような場合、そういった局所的な値に対しても個別的なモデリングが可能であることから、学習中の損失関数は低下しやすい可能性がある。しかしながら、そういった挙動が汎化性能にどのように影響するかは不明瞭である。たとえば Yu et al. (2024) においても類似した問題意識から KAN の定義域を広げた結果が解析されているが、本研究では逆に KAN の入力を標準化することによる効果を検証する。以下に示す簡易的な 4 手法を採用し、関数全体での定義域を $[0, 1]$ に収める。まず、ある変数 x を Min–Max 正規化した標準化スコアを

$$z^{(MM)} = \frac{x - \min(x)}{\max(x) - \min(x)} \in [0, 1] \quad (16)$$

とする。Min-Max 正規化においては外れ値の影響を強く受けてその他の情報が圧縮されてしまう。次に、Sigmoid 関数により $[0, 1]$ に標準化した $z^{(Sigm)}$ を

$$z^{(Sigm)} = \frac{1}{1 + e^{-x}} \in [0, 1] \quad (17)$$

とする。しかしながら、Sigmoid 関数では Min-Max 正規化とは対照的に比較的大きな／小さな値ほど強く圧縮されてしまい、つまりいずれの手法においても特徴量の識別性が低下する。そこで、他の顧客との相対的な購買順位を特徴量とするべく、次に、Percentile を用いた標準化を行う。サンプルサイズ n の変数 x について、順序データを $x = (x_{(1)}, x_{(2)}, \dots, x_{(n)})$ とすると、前から r 番目に位置する $x_{(r)}$ についてパーセンタイル標準化スコア $z^{(Pct)}$ を

$$z^{(Pct)} = \frac{r}{n} \in [0, 1] \quad (18)$$

と表現できる ($r = \{1, 2, \dots, n\}$)。さらに、切断を伴う標準化 (clipping normalization, CN) として、特定のスケールでの標準化を行った上で、特定の閾値で外れ値を切断する。本研究では、ある変数 x を平均 0.5、標準偏差 $1/6$ に標準化することにより、その標準偏差 σ について 3σ を超える外れ値を下限 0、上限 1 で切断した $z^{(CN)}$ を

$$z^{(CN)} = \max(0, \min(1, \frac{x - \bar{x}}{6\sigma} + 0.5)) \in [0, 1] \quad (19)$$

とする。これにより、 3σ を超える外れ値を $[0, 1]$ に収める。

4.3 解析結果 1：モデル比較

全条件の探索を通じた網羅的な解析から、汎化誤差の高さによる上位 20 モデルのハイパーパラメーターと予測精度を表 4 に示す。まずこの結果からも明らかなように、モデルの深層化 (L の増加) や複雑化 (各 l における J_l の増加) はモデルの予測性能にほとんど貢献していない。特に、隠れ層を 3 層設定した探索条件などもあるなかで、上位モデルのほとんどは隠れ層が 2 つ程度となっている。最適化手法については、SGD を除く他の 3 手法が大きな偏りなく含まれている。Scheduler についても満遍なく含まれており、上位モデルのみでは明確な傾向は見出せない。

一般的な深層学習においては、複雑なアーキテクチャーにより学習データに強くフィッティングさせた場合、その分だけ過適合 (overfitting) により汎化性能が低下する。KAN に関して、ノイズの多い関数への回帰に KAN を応用した研究 (Zeng et al., 2024) では、データ上の小さなノイズでもテスト性能を悪化させることが指摘されている。しかしながら、少なくとも本研究の解析結果に限って言えば、train loss では上位 20 モデルのうち参照モデルを上回るものが一つもない一方で、そのすべてが test loss では上回っている。これは参照モデルと比較してその乖離幅が平均的に小さいことを意味している。つまり、適切な構造を設定することにより、過適合を防ぎながら汎化性能を高めることが可能であり、これが KAN では最適なモデル構造の探索が重要 (Liu et al., 2024) と指摘される所以であると推察される。

次に、表 5 に提案モデル内でのハイパーパラメーターの設定による予測精度のグループ内平均と最小値を示す。全体的に確認できるのは、モデルのアーキテクチャーが過剰に複雑になることにより最良モデルの予測性能が悪化していく点である。例として、スプライン関数の次元数 d では $d = 2$ の場合から、あるいは隠れ層のアーキテクチャーについては 1 層 3 ノードの場合から始まり、モデルが複雑になると並行して test loss が悪化していく傾向が確認できる。そして探索モデル全体では Scheduler を用いることで平均的には汎化誤差が改善している。

また、本研究の関心でもある標準化による定義域の限定の効果についても確認すると、いずれの標準化も行わない場合に最良の結果が得られる結果となった。この項目については特に、他のパラメーターと異なり標準化を行わなかった場合のパフォーマンスの改善幅が大きくなっている。これはグリッド数の増加が予測精度の改善につながることも関係していると推察される。すでに 4.2 節でも述べたとおり、前処理の段階での標準化による定義域の限定は、その多くの手法で特徴量の情報損失につながる。つまり KAN においては、次元数 d や隠れ層の規模 (L, J_l) の安易な増大により過剰に高次元な関数表現を志向するよりも、区間数 K の増加により関数形を緻密にコントロールすることが予測精度に貢献することを意味している。

4.4 解析結果 2：モデルの可視化

KAN では単変数での活性化ののちに総和を取るため、ノード間の変数の関係性の解釈性が高いとされている。そこで、先の解析において最良の精度となった表 4 の最上段のモデルについて簡易的な可視化を行う。最良モデルでは入力層の大きさ $J_0 = 5$ より、各ノード間の活性化関数の $[-1, 1]$ にわたる構造と最終的なモデル出力を図 5 に示す。ここで $K = 50$ 、 $J_1 = 3$ 、 $d = 2$ より、定義域全体にわたって区間を 50 に分割し、3

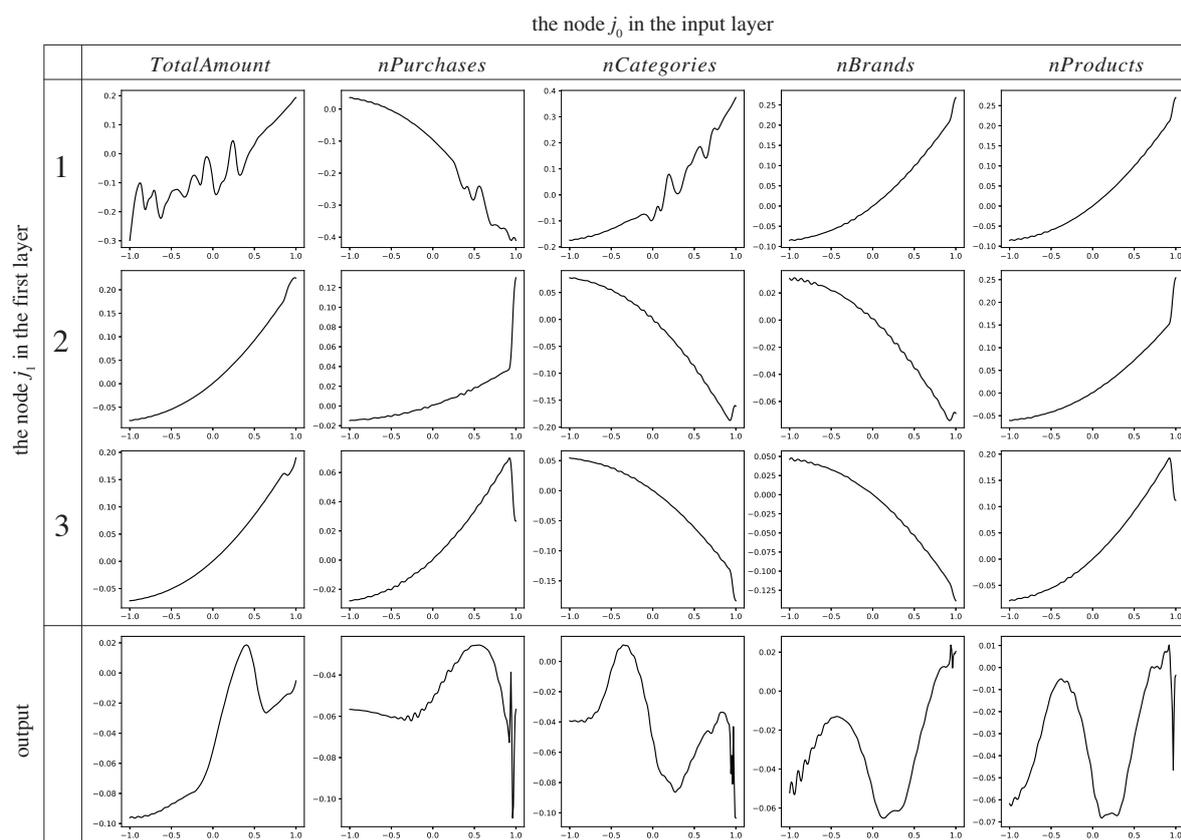


図 5 The first activations and final output of each input

つの 2 次関数の総和により、非線形的な関係性が得られていることがわかる^{*10}。具体的にその関数形を確認すると、指数関数に近い増加関係にありながら、値が大きくなった場合にのみ大きく異なる反応を返すような

^{*10} 実際には $l = 2$ として中間に別のノード $h_{2,1}$ が存在するため、図 5 上段の隠れ層の総和は出力の関数形と一致しない。

動きが複数確認でき (e.g., $nPurchases$ や $nProducts$ からノード 3 への活性化)、先に述べたような関数の局所的なコントロールが可能なスプライン関数の利点が活かされていることがわかる。一方で、局所的にトレーニングデータへの過適合が発生していることが懸念される箇所も複数確認できる (e.g., $TotalAmount$ や $nCategories$ からノード 1 への活性化)。このように、確かに変数の解釈性については各層である程度の関係性が確認できる。ただし、一般的な MLP と同様に多層的に複数の非線形関数の合成が行われるため、それら中間ノードにマーケティング的な意味づけを行うことは必ずしも容易ではない。これらの点からも、深層学習とはいえモデルのパラメーターの大規模化は結果的に KAN の利点を損なう可能性がある。

5 まとめ

本研究では深層学習分野で新たに提案されたモデルである KAN について、その計算構造を従来の MLP と比較する形で検討するとともに、実データを用いてモデル探索を行い、複数の参照モデルと比較してその有用性を示した。ここまで述べてきた結果から、本研究の解析上で KAN は MLP や既存の機械学習手法を上回る精度での予測が可能であることが示された。また、本研究ではモデルの定義域を限定することによる効果についていくつかの非常に単純な手法を用いて比較したが、結果としてほとんどの場合において正規化を行わない方が高い精度を示す結果となった。従来一般的に用いられてきた Min-Max 正規化等はあくまでも学習時点で得られているデータセット内での最大/最小値に依拠して標準化を行う手法であり、将来期間にわたって取得される未知のデータについて、あらゆる特徴量が学習時の値を超えないことは確率的に期待できない。特に、マーケティング領域で一般的に用いられる購買の頻度や金額等は、多くの場合に正の整数値である。その定義域には明確な上限やスケールが存在せず、特定範囲での標準化は伝統的に困難である。そのような状況にあって、KAN では特徴量の標準化を行わなくとも高精度の予測が可能であることは、マーケティングに限らず幅広い社会科学分野において有用となる可能性がある。

また、KAN では変数の解釈性を高められることが主張されているが、本研究の可視化プロセスからは、これまで状態空間モデル等を用いることにより説明変数のパラメーター側の動的変化の把握を志向したのと同様に、KAN の学習可能な活性化関数は観測可能な行動的ロイヤルティの値によるアウトカムへの反応のダイナミクスを記述する過程としても理解できる可能性がある。

まとめると、KAN は KART そのものの純粋な拡張であるかどうかとは全く無関係に、その定理の構造に着想を得て、非線形関数の多層的な合成による柔軟な近似手法を深層学習構造に組み込んだ点に大きな新奇性がある。事実、KAN の複数の発展モデル (e.g., Bozorgasl and Chen, 2024; Sidharth, 2024) においてスプライン関数が他手法に置換されていることから明らかなように、KART の実証モデル化にはスプライン関数以外にも複数のアプローチが考えられる。それでも、深層学習の発展にあたって新しい構造を提示したという意味で示唆的であることは疑いようもない。

最後に、本研究の課題について述べると、第一にはより発展的なモデルの採用がある。本研究の分析においてはあくまでも最も基本的な KAN のモデルのみを用いたが、KAN をベースにした手法は次々に提案されている。それらを用いることによる変化についても今後確認する必要がある。第二に、深層学習の参照モデルとしては単純な MLP のみを採用したが、Transformer やマルチモーダル学習等のより発展的な手法も取り入れる必要がある。もちろん、本研究では単純な予測性能の比較のみに留まらない比較検討を行ったが、モデルの純粋な予測性能は多くの場合に state-of-the-art (SOTA) 手法との比較によって初めて理解可能である。第三に、著者らが知る限り本研究は KAN をマーケティングに初めて応用した研究である。その有用性の検証にあたってはより包括的かつ多方面からのアプローチが求められる。

謝辞

本研究の実施にあたり、積極的にオープンデータを公開し学術・教育へ貢献する REES46 に感謝を申し上げます。本研究は JSPS 科学研究費 (24K16472) および名城大学 経済・経営学会の研究助成を受けている。

参考文献

- Akashi, Shigeo (2001) “Application of ϵ -entropy theory to Kolmogorov—Arnold representation theorem,” *Reports on Mathematical Physics*, Vol. 48, No. 1-2, pp. 19–26, DOI: [http://dx.doi.org/10.1016/S0034-4877\(01\)80060-4](http://dx.doi.org/10.1016/S0034-4877(01)80060-4).
- Ala'raj, Maher, Maysam F Abbod, and Munir Majdalawieh (2021) “Modelling customers credit card behaviour using bidirectional LSTM neural networks,” *Journal of Big Data*, Vol. 8, No. 1, pp. 1–27, DOI: <http://dx.doi.org/10.1186/s40537-021-00461-7>.
- Altarabichi, Mohammed Ghaith (2024) “DropKAN: Regularizing KANs by masking post-activations,” *arXiv preprint arXiv:2407.13044*, DOI: <http://dx.doi.org/10.48550/arXiv.2407.13044>.
- Arnol'd, Vladimir I. (1957) “On functions of three variables,” *Proceedings of the USSR Academy of Sciences*, Vol. 114, pp. 679–681 (English translation: Amer. Math. Soc. Transl., 28 (1963), pp. 51–54.), DOI: http://dx.doi.org/10.1007/978-3-642-01742-1_2.
- Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio (2015) “Neural machine translation by jointly learning to align and translate,” *3rd International Conference on Learning Representations (ICLR 2015)*, DOI: <http://dx.doi.org/10.48550/arXiv.1409.0473>.
- Bengio, Yoshua, Pascal Lamblin, Dan Popovici, and Hugo Larochelle (2006) “Greedy layer-wise training of deep networks,” *Proceedings of the 19th International Conference on Neural Information Processing Systems (NIPS2006)*, Vol. 19.
- Bengio, Yoshua, Ian Goodfellow, and Aaron Courville (2017) *Deep learning*, Vol. 1: MIT press Cambridge, MA, USA, DOI: <http://dx.doi.org/10.1007/s10710-017-9314-z>.
- Bodner, Alexander Dylan, Antonio Santiago Tepsich, Jack Natan Spolski, and Santiago Pourteau (2024) “Convolutional Kolmogorov-Arnold Networks,” *arXiv preprint arXiv:2406.13155*, DOI: <http://dx.doi.org/10.48550/arXiv.2406.13155>.
- Bottou, Léon (2012) “Stochastic gradient descent tricks,” in *Neural networks: Tricks of the trade*: Springer, pp. 421–436.
- Bozorgasl, Zavareh and Hao Chen (2024) “Wav-kan: Wavelet kolmogorov-arnold networks,” *arXiv preprint arXiv:2405.12832*, DOI: <http://dx.doi.org/10.48550/arXiv.2405.12832>.
- Bresson, Roman, Giannis Nikolentzos, George Panagopoulos, Michail Chatzianastasis, Jun Pang, and Michalis Vazirgiannis (2024) “Kagnns: Kolmogorov-arnold networks meet graph learning,” *arXiv preprint arXiv:2406.18380*, DOI: <http://dx.doi.org/10.48550/arXiv.2406.18380>.
- Brown, Tom B (2020) “Language models are few-shot learners,” *arXiv preprint ArXiv:2005.14165*, DOI: <http://dx.doi.org/10.48550/arXiv.2005.14165>.
- Chen, Pei Pei, Anna Guitart, Ana Fernández del Río, and Africa Perriñez (2018) “Customer lifetime value

- in video games using deep learning and parametric models,” in *2018 IEEE international conference on big data (big data)*, pp. 2134–2140, IEEE, DOI: <http://dx.doi.org/10.1109/BigData.2018.8622151>.
- Chen, Tianqi and Carlos Guestrin (2016) “Xgboost: A scalable tree boosting system,” in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794, DOI: <http://dx.doi.org/10.1145/2939672.2939785>.
- Cheon, Minjong (2024) “Kolmogorov-Arnold Network for Satellite Image Classification in Remote Sensing,” *arXiv preprint arXiv:2406.00600*, DOI: <http://dx.doi.org/10.48550/arXiv.2406.00600>.
- Cho, Kyunghyun, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio (2014) “Learning phrase representations using RNN encoder-decoder for statistical machine translation,” *arXiv preprint arXiv:1406.1078*, DOI: <http://dx.doi.org/10.3115/v1/D14-1179>.
- Choi, Woo Yong, Kyu Ye Song, and Chan Woo Lee (2018) “Convolutional attention networks for multi-modal emotion recognition from speech and text data,” in *Proceedings of grand challenge and workshop on human multimodal language (Challenge-HML)*, pp. 28–34, DOI: <http://dx.doi.org/10.18653/v1/W18-3304>.
- Cortes, Corinna and Vladimir Vapnik (1995) “Support-vector networks,” *Machine learning*, Vol. 20, pp. 273–297, DOI: <http://dx.doi.org/10.1007/BF00994018>.
- Cover, Thomas and Peter Hart (1967) “Nearest neighbor pattern classification,” *IEEE transactions on information theory*, Vol. 13, No. 1, pp. 21–27, DOI: <http://dx.doi.org/10.1109/TIT.1967.1053964>.
- De Boor, Carl (2003) *A practical guide to splines*: springer New York, revised edition.
- De Caigny, Arno, Kristof Coussement, Koen W De Bock, and Stefan Lessmann (2020) “Incorporating textual information in customer churn prediction models based on a convolutional neural network,” *International Journal of Forecasting*, Vol. 36, No. 4, pp. 1563–1578, DOI: <http://dx.doi.org/10.1016/j.ijforecast.2019.03.029>.
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2019) “Bert: Pre-training of deep bidirectional transformers for language understanding,” *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), Volume 1 (Long and Short Papers)*, pp. 4171–4186, DOI: <http://dx.doi.org/10.18653/v1/N19-1423>.
- Drokin, Ivan (2024) “Kolmogorov-Arnold Convolutions: Design Principles and Empirical Studies,” *arXiv preprint arXiv:2407.01092*, DOI: <http://dx.doi.org/10.48550/arXiv.2407.01092>.
- Elfwing, Stefan, Eiji Uchibe, and Kenji Doya (2018) “Sigmoid-weighted linear units for neural network function approximation in reinforcement learning,” *Neural networks*, Vol. 107, pp. 3–11, DOI: <http://dx.doi.org/10.1016/j.neunet.2017.12.012>.
- Elman, Jeffrey L (1990) “Finding structure in time,” *Cognitive science*, Vol. 14, No. 2, pp. 179–211, DOI: http://dx.doi.org/10.1207/s15516709cog1402_1.
- Genet, Remi and Hugo Inzirillo (2024) “Tkan: Temporal kolmogorov-arnold networks,” *arXiv preprint arXiv:2405.07344*, DOI: <http://dx.doi.org/10.48550/arXiv.2405.07344>.
- Gilardi, Fabrizio, Meysam Alizadeh, and Maël Kubli (2023) “ChatGPT outperforms crowd workers

- for text-annotation tasks,” *Proceedings of the National Academy of Sciences*, Vol. 120, No. 30, p. e2305016120, DOI: <http://dx.doi.org/10.1073/pnas.2305016120>.
- Graves, Alex and Jürgen Schmidhuber (2005) “Frameworkwise phoneme classification with bidirectional LSTM and other neural network architectures,” *Neural networks*, Vol. 18, No. 5-6, pp. 602–610, DOI: <http://dx.doi.org/10.1016/j.neunet.2005.06.042>.
- Guo, Long, Lifeng Hua, Rongfei Jia, Binqiang Zhao, Xiaobo Wang, and Bin Cui (2019) “Buying or browsing?: Predicting real-time purchasing intent using attention-based deep network with multiple behavior,” in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 1984–1992, DOI: <http://dx.doi.org/10.1145/3292500.3330670>.
- He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun (2016) “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, DOI: <http://dx.doi.org/10.1109/CVPR.2016.90>.
- Hinton, Geoffrey E and Ruslan R Salakhutdinov (2006) “Reducing the dimensionality of data with neural networks,” *science*, Vol. 313, No. 5786, pp. 504–507, DOI: <http://dx.doi.org/10.1126/science.1127647>.
- Hinton, Geoffrey E, Simon Osindero, and Yee-Whye Teh (2006) “A fast learning algorithm for deep belief nets,” *Neural computation*, Vol. 18, No. 7, pp. 1527–1554, DOI: <http://dx.doi.org/10.1162/neco.2006.18.7.1527>.
- Ho, Jonathan, Ajay Jain, and Pieter Abbeel (2020) “Denoising diffusion probabilistic models,” *Advances in neural information processing systems*, Vol. 33, pp. 6840–6851.
- Hochreiter, Sepp and Jürgen Schmidhuber (1997) “Long short-term memory,” *Neural computation*, Vol. 9, No. 8, pp. 1735–1780, DOI: <http://dx.doi.org/10.1162/neco.1997.9.8.1735>.
- Hutto, Clayton and Eric Gilbert (2014) “Vader: A parsimonious rule-based model for sentiment analysis of social media text,” in *Proceedings of the international AAAI conference on web and social media*, Vol. 8, pp. 216–225, DOI: <http://dx.doi.org/10.1609/icwsm.v8i1.14550>.
- Jacoby, Jacob and Robert W Chestnut (1978) *Brand loyalty: Measurement and management*: John Wiley & Sons Incorporated.
- Joachims, Thorsten (1998) “Text categorization with support vector machines: Learning with many relevant features,” in *European conference on machine learning*, pp. 137–142, Springer, DOI: <http://dx.doi.org/10.1007/BFb0026683>.
- Jordan, Michael I (1997) “Serial order: A parallel distributed processing approach,” in *Advances in psychology*, Vol. 121: Elsevier, pp. 471–495, DOI: [http://dx.doi.org/10.1016/S0166-4115\(97\)80111-2](http://dx.doi.org/10.1016/S0166-4115(97)80111-2).
- Kamakura, Wagner A and Michel Wedel (1997) “Statistical data fusion for cross-tabulation,” *Journal of Marketing Research*, Vol. 34, No. 4, pp. 485–498, DOI: <http://dx.doi.org/10.1177/002224379703400406>.
- Ke, Guolin, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu (2017) “Lightgbm: A highly efficient gradient boosting decision tree,” *Advances in neural information processing systems (NIPS2017)*, Vol. 30.
- Khattar, Anuradha and SMK Quadri (2022) “Camm: Cross-attention multimodal classification of

- disaster-related tweets,” *IEEE Access*, Vol. 10, pp. 92889–92902, DOI: <http://dx.doi.org/10.1109/ACCESS.2022.3202976>.
- Kiamari, Mehrdad, Mohammad Kiamari, and Bhaskar Krishnamachari (2024) “GKAN: Graph Kolmogorov-Arnold Networks,” *arXiv preprint arXiv:2406.06470*, DOI: <http://dx.doi.org/10.48550/arXiv.2406.06470>.
- Kim, Minsu, Woosik Shin, SeongBeom Kim, and Hee-Woong Kim (2023) “Predicting Session Conversion on E-commerce: A Deep Learning-based Multimodal Fusion Approach,” *Asia pacific journal of information systems*, Vol. 33, No. 3, pp. 737–767, DOI: <http://dx.doi.org/10.14329/apjis.2023.33.3.737>.
- Kingma, Diederik P and Jimmy Ba (2014) “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*.
- Koehn, Dennis, Stefan Lessmann, and Markus Schaal (2020) “Predicting online shopping behaviour from clickstream data using deep learning,” *Expert Systems with Applications*, Vol. 150, p. 113342, DOI: <http://dx.doi.org/10.1016/j.eswa.2020.113342>.
- Kolmogorov, A.N. (1956) “On the representation of continuous functions of several variables by superpositions of continuous functions of a smaller number of variables,” *Proceedings of the USSR Academy of Sciences*, Vol. 108, pp. 179–182 (English translation: Amer. Math. Soc. Transl., 17 (1961), pp. 369–373.), DOI: http://dx.doi.org/10.1007/978-3-642-01742-1_5.
- Köppen, Mario (2002) “On the training of a Kolmogorov Network,” in *Artificial Neural Networks—ICANN 2002: International Conference Madrid, Spain, August 28–30, 2002 Proceedings 12*, pp. 474–479, Springer, DOI: http://dx.doi.org/10.1007/3-540-46084-5_77.
- Krugmann, Jan Ole and Jochen Hartmann (2024) “Sentiment Analysis in the Age of Generative AI,” *Customer Needs and Solutions*, Vol. 11, No. 1, p. 3, DOI: <http://dx.doi.org/10.1007/s40547-024-00143-4>.
- Kundu, Akash, Aritra Sarkar, and Abhishek Sadhu (2024) “Kanqas: Kolmogorov arnold network for quantum architecture search,” *arXiv preprint arXiv:2406.17630*, DOI: <http://dx.doi.org/10.48550/arXiv.2406.17630>.
- Larochelle, Hugo and Yoshua Bengio (2008) “Classification using discriminative restricted Boltzmann machines,” in *Proceedings of the 25th international conference on Machine learning*, pp. 536–543, DOI: <http://dx.doi.org/10.1145/1390156.1390224>.
- Le, Quoc and Tomas Mikolov (2014) “Distributed representations of sentences and documents,” in *International conference on machine learning (ICML2014)*, pp. 1188–1196.
- LeCun, Yann, Yoshua Bengio et al. (1995) “Convolutional networks for images, speech, and time series,” *The handbook of brain theory and neural networks*, Vol. 3361, No. 10, p. 1995.
- Lewis, Patrick, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel et al. (2020) “Retrieval-augmented generation for knowledge-intensive nlp tasks,” *Advances in Neural Information Processing Systems (NIPS2020)*, Vol. 33, pp. 9459–9474.
- Lin, Ji-Nan and Rolf Unbehauen (1993) “On the realization of a Kolmogorov network,” *Neural Computation*, Vol. 5, No. 1, pp. 18–20, DOI: <http://dx.doi.org/10.1162/neco.1993.5.1.18>.

- Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov (2019) “Roberta: A robustly optimized bert pretraining approach,” *arXiv preprint arXiv:1907.11692*, DOI: <http://dx.doi.org/10.48550/arXiv.1907.11692>.
- Liu, Ziming, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson, Marin Soljačić, Thomas Y Hou, and Max Tegmark (2024) “Kan: Kolmogorov-arnold networks,” *arXiv preprint arXiv:2404.19756*, DOI: <http://dx.doi.org/10.48550/arXiv.2404.19756>.
- Lobanov, Valeriy, Nikita Firsov, Evgeny Myasnikov, Roman Khabibullin, and Artem Nikonorov (2024) “HyperKAN: Kolmogorov-Arnold Networks make Hyperspectral Image Classifiers Smarter,” *arXiv preprint arXiv:2407.05278*, DOI: <http://dx.doi.org/10.48550/arXiv.2407.05278>.
- Lorentz, GG (1962) “Metric entropy, widths, and superpositions of functions,” *The American Mathematical Monthly*, Vol. 69, No. 6, pp. 469–485, DOI: <http://dx.doi.org/10.1080/00029890.1962.11989915>.
- Loshchilov, I (2017) “Decoupled weight decay regularization,” *arXiv preprint arXiv:1711.05101*.
- Ma, Liye and Baohong Sun (2020) “Machine learning and AI in marketing—Connecting computing power to human insights,” *International Journal of Research in Marketing*, Vol. 37, No. 3, pp. 481–504, DOI: <http://dx.doi.org/10.1016/j.ijresmar.2020.04.005>.
- Mena, C Gary, Arno De Caigny, Kristof Coussemont, Koen W De Bock, and Stefan Lessmann (2019) “Churn Prediction with Sequential Data and Deep Neural Networks. A Comparative Analysis,” *arXiv preprint arXiv:1909.11114*, DOI: <http://dx.doi.org/10.48550/arXiv.1909.11114>.
- Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean (2013) “Efficient estimation of word representations in vector space,” *arXiv preprint arXiv:1301.3781*, DOI: <http://dx.doi.org/10.48550/arXiv.1301.3781>.
- Ngai, Eric WT and Yuanyuan Wu (2022) “Machine learning in marketing: A literature review, conceptual framework, and research agenda,” *Journal of Business Research*, Vol. 145, pp. 35–48, DOI: <http://dx.doi.org/10.1016/j.jbusres.2022.02.049>.
- Ngiam, Jiquan, Aditya Khosla, Mingyu Kim, Juhan Nam, Honglak Lee, and Andrew Y Ng (2011) “Multimodal deep learning,” in *Proceedings of the 28th international conference on machine learning (ICML-11)*, pp. 689–696.
- Niimi, Junichiro (2024a) “Dynamic Sentiment Analysis with Local Large Language Models using Majority Voting: A Study on Factors Affecting Restaurant Evaluation,” *arXiv preprint arXiv:2407.13069*, DOI: <http://dx.doi.org/10.48550/arXiv.2407.13069>.
- (2024b) “An Efficient Multimodal Learning Framework to Comprehend Consumer Preferences Using BERT and Cross-Attention,” *arXiv preprint arXiv:2405.07435*, DOI: <http://dx.doi.org/10.48550/arXiv.2405.07435>.
- Ostrand, Phillip A (1965) “Dimension of metric spaces and Hilbert’s problem 13.”
- Pan, Hong and Hanxun Zhou (2020) “Study on convolutional neural network and its application in data mining and sales forecasting for E-commerce,” *Electronic Commerce Research*, Vol. 20, No. 2, pp. 297–320, DOI: <http://dx.doi.org/10.1007/s10660-020-09409-0>.
- Pan, Ziyang, Zhishan Huang, Xiaowen Lin, Songxia Li, Huanze Zeng, and Daifeng Li (2020) “Multi-data Fusion Based Marketing Prediction of Listed Enterprise Using MS-LSTM Model,” in *Proceedings of*

- the 2020 3rd International Conference on Algorithms, Computing and Artificial Intelligence, pp. 1–10, DOI: <http://dx.doi.org/10.1145/3446132.3446169>.
- Peters, Matthew E., Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer (2018) “Deep Contextualized Word Representations,” *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pp. 2227–2237, DOI: <http://dx.doi.org/10.18653/v1/N18-1202>.
- Poggio, Tomaso, Andrzej Banburski, and Qianli Liao (2020) “Theoretical issues in deep networks,” *Proceedings of the National Academy of Sciences*, Vol. 117, No. 48, pp. 30039–30045, DOI: <http://dx.doi.org/10.1073/pnas.1907369117>.
- Polar, Andrew and Michael Poluektov (2021) “A deep machine learning algorithm for construction of the Kolmogorov–Arnold representation,” *Engineering Applications of Artificial Intelligence*, Vol. 99, p. 104137, DOI: <http://dx.doi.org/10.1016/j.engappai.2020.104137>.
- Prechelt, Lutz (1998) “Early stopping-but when?” in *Neural Networks: Tricks of the trade*: Springer, pp. 55–69, DOI: http://dx.doi.org/10.1007/978-3-642-35289-8_5.
- Ramachandram, Dhanesh and Graham W Taylor (2017) “Deep multimodal learning: A survey on recent advances and trends,” *IEEE signal processing magazine*, Vol. 34, No. 6, pp. 96–108, DOI: <http://dx.doi.org/10.1109/MSP.2017.2738401>.
- REES46 Marketing Platform (2020) *eCommerce purchase history from electronics store*: (M. Kechinov, <https://www.kaggle.com/datasets/mkechinov/ecommerce-purchase-history-from-electronics-store>, accessed Apr. 20th, 2024).
- Rumelhart, David E, Geoffrey E Hinton, and Ronald J Williams (1986a) *Learning internal representations by error propagation, In parallel distributed processing, explorations in the microstructure of cognition, Vol. 1: Foundations*, pp. 318-362: MIT Press.
- (1986b) “Learning representations by back-propagating errors,” *nature*, Vol. 323, No. 6088, pp. 533–536.
- Salakhutdinov, Ruslan and Geoffrey Hinton (2009) “Deep boltzmann machines,” in *12th International Conference on Artificial intelligence and statistics (AISTATS) 2009*, pp. 448–455, PMLR.
- Sanh, Victor, Lysandre Debut, Julien Chaumond, and Thomas Wolf (2019) “DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter,” *arXiv*, DOI: <http://dx.doi.org/10.48550/arxiv.1910.01108>.
- Schmidt-Hieber, Johannes (2021) “The Kolmogorov–Arnold representation theorem revisited,” *Neural networks*, Vol. 137, pp. 119–126, DOI: <http://dx.doi.org/10.1016/j.neunet.2021.01.020>.
- Sidharth, SS (2024) “Chebyshev polynomial-based kolmogorov-arnold networks: An efficient architecture for nonlinear function approximation,” *arXiv preprint arXiv:2405.07200*, DOI: <http://dx.doi.org/10.48550/arXiv.2405.07200>.
- Simonyan, Karen and Andrew Zisserman (2014) “Very deep convolutional networks for large-scale image recognition,” DOI: <http://dx.doi.org/10.48550/arXiv.1409.1556>, arXiv preprint arXiv:1409.1556.
- Smola, Alex J and Bernhard Schölkopf (2004) “A tutorial on support vector regression,” *Statistics and*

- computing*, Vol. 14, pp. 199–222, DOI: <http://dx.doi.org/10.1023/B:STC0.0000035301.49549.88>.
- Smolensky, Paul et al. (1986) “Information processing in dynamical systems: Foundations of harmony theory.”
- Sparck Jones, Karen (1972) “A statistical interpretation of term specificity and its application in retrieval,” *Journal of documentation*, Vol. 28, No. 1, pp. 11–21, DOI: <http://dx.doi.org/10.1108/eb026526>.
- Sprecher, David A (1965) “On the structure of continuous functions of several variables,” *Transactions of the American Mathematical Society*, Vol. 115, pp. 340–355, DOI: <http://dx.doi.org/10.2307/1994273>.
- Srivastava, Nitish and Russ R Salakhutdinov (2012) “Multimodal learning with deep boltzmann machines,” *Advances in neural information processing systems*, Vol. 25.
- Szegedy, Christian, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich (2015) “Going deeper with convolutions,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, DOI: <http://dx.doi.org/10.1109/CVPR.2015.7298594>.
- Toth, Arthur, Louis Tan, Giuseppe Di Fabrizio, and Ankur Datta (2017) “Predicting shopping behavior with mixture of RNNs,” in *eCOM@ SIGIR*.
- Vaca-Rubio, Cristian J, Luis Blanco, Roberto Pereira, and Màrius Caus (2024) “Kolmogorov-arnold networks (kans) for time series analysis,” *arXiv preprint arXiv:2405.08790*, DOI: <http://dx.doi.org/10.48550/arXiv.2405.08790>.
- Van Atteveldt, Wouter, Mariken ACG Van der Velden, and Mark Boukes (2021) “The validity of sentiment analysis: Comparing manual annotation, crowd-coding, dictionary approaches, and machine learning algorithms,” *Communication Methods and Measures*, Vol. 15, No. 2, pp. 121–140, DOI: <http://dx.doi.org/10.1080/19312458.2020.1869198>.
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin (2017) “Attention is all you need,” *Advances in neural information processing systems*, Vol. 30, pp. 5998–6008.
- Volkmar, Gioia, Peter M Fischer, and Sven Reinecke (2022) “Artificial Intelligence and Machine Learning: Exploring drivers, barriers, and future developments in marketing management,” *Journal of Business Research*, Vol. 149, pp. 599–614, DOI: <http://dx.doi.org/10.1016/j.jbusres.2022.04.007>.
- Wang, Zengzhi, Qiming Xie, Yi Feng, Zixiang Ding, Zinong Yang, and Rui Xia (2023) “Is ChatGPT a good sentiment analyzer? A preliminary study,” *arXiv preprint arXiv:2304.04339*, DOI: <http://dx.doi.org/10.48550/arXiv.2304.04339>.
- Xu, Kunpeng, Lifei Chen, and Shengrui Wang (2024) “Kolmogorov-Arnold Networks for Time Series: Bridging Predictive Power and Interpretability,” *arXiv preprint arXiv:2406.02496*, DOI: <http://dx.doi.org/10.48550/arXiv.2406.02496>.
- Yadav, Ashima and Dinesh Kumar Vishwakarma (2020) “Sentiment analysis using deep learning architectures: a review,” *Artificial Intelligence Review*, Vol. 53, No. 6, pp. 4335–4385, DOI: <http://dx.doi.org/10.1007/s10462-019-09794-5>.
- Yoon, Seunghyun, Seokhyun Byun, Subhadeep Dey, and Kyomin Jung (2019) “Speech emotion recogni-

- tion using multi-hop attention mechanism,” in *ICASSP 2019-2019 IEEE International conference on acoustics, speech and signal processing (ICASSP)*, pp. 2822–2826, IEEE, DOI: <http://dx.doi.org/10.1109/ICASSP.2019.8683483>.
- Yu, Runpeng, Weihao Yu, and Xinchao Wang (2024) “Kan or mlp: A fairer comparison,” *arXiv preprint arXiv:2407.16674*, DOI: <http://dx.doi.org/10.48550/arXiv.2407.16674>.
- Zadeh, Amir, Paul Pu Liang, Navonil Mazumder, Soujanya Poria, Erik Cambria, and Louis-Philippe Morency (2018) “Memory fusion network for multi-view sequential learning,” in *Proceedings of the AAAI conference on artificial intelligence*, Vol. 32.
- Zeng, Chen, Jiahui Wang, Haoran Shen, and Qiao Wang (2024) “KAN versus MLP on Irregular or Noisy Functions,” *arXiv preprint arXiv:2408.07906*, DOI: <http://dx.doi.org/10.48550/arXiv.2408.07906>.
- 新美潤一郎 (2021) 「時間単位 Clumpiness 指標を用いた解析手法の提案: 生存時間分析と LSTM-RNN を用いた利用頻度・離脱時間の予測への RFMC 分析の活用」, 『名城論叢』, 第 22 巻, 第 2 号, 49–63 頁.
- (2024) 「異なる次元数のデータを同時に投入した行動的ロイヤルティ推計手法の提案—Source-Attention Transformer と特徴融合によるマルチモーダル深層学習—」, 『応用統計学』, 第 53 巻, 第 1 号, 15–32 頁, DOI: <http://dx.doi.org/10.5023/jappstat.53.15>.
- 新美潤一郎・星野崇宏 (2017a) 「Deep Boltzmann Machine を用いたデータ融合手法の提案」, 『人工知能学会全国大会論文集 第 31 回 (2017)』, 1I12–1I12 頁, 一般社団法人 人工知能学会, DOI: http://dx.doi.org/10.11517/pjsai.JSAI2017.0_1I12.
- (2017b) 「顧客行動の多様性変数を利用した購買行動の予測」, 『人工知能学会論文誌』, 第 32 巻, 第 2 号, B-G63.1–9 頁, DOI: <http://dx.doi.org/10.1527/tjsai.B-G63>.

表 4 Results (Top 20 Models in RMSE, Ascending in Test Loss)

K	d	Hidden Layers	Optimizer	Scheduler	Norm.	Train	Val	Test	
Proposed Models (KAN)									
50	2	(3,1)	Adamax	Linear	None	0.715	0.790	1.223	
20	4	(50,1)	AdamW	None	None	0.670	0.807	1.226	
10	3	(10,1)	Adam	None	None	0.687	0.780	1.226	
10	4	(10,5,1)	Adam	Linear	None	0.717	0.784	1.226	
50	5	(50,1)	Adamax	None	None	1.083	0.789	1.227	
20	3	(10,5,1)	AdamW	None	None	0.710	0.786	1.227	
50	5	(50,1)	AdamW	Linear	None	0.958	0.803	1.228	
10	1	(10,5,1)	Adam	Exponential	None	0.713	0.793	1.229	
10	1	(50,1)	Adamax	None	None	0.682	0.796	1.229	
20	5	(50,1)	AdamW	Exponential	None	0.713	0.812	1.229	
50	2	(50,1)	Adam	Linear	None	1.249	0.797	1.229	
20	4	(5,1)	Adamax	Linear	None	0.715	0.790	1.229	
50	1	(5,3,1)	Adam	Linear	None	2.234	0.786	1.229	
50	3	(10,5,1)	Adam	Exponential	None	0.695	0.790	1.229	
20	3	(5,3,1)	Adam	None	None	0.697	0.780	1.229	
10	3	(5,3,1)	Adam	None	None	0.692	0.779	1.229	
10	3	(20,10,5,1)	Adam	None	None	0.707	0.789	1.230	
50	1	(10,5,1)	AdamW	Exponential	None	0.687	0.793	1.230	
10	1	(50,20,10,1)	Adamax	Linear	None	0.998	0.778	1.230	
10	5	(10,1)	Adam	Exponential	None	0.689	0.781	1.230	
						Mean	0.701	0.779	1.260
						Min	0.596	0.746	1.223
						Median	0.687	0.771	1.259
						Max	2.234	5.005	1.282
Reference Models									
LightGBM						0.832	0.802	1.256	
XGBoost						0.649	0.890	1.277	
MLP (tanh)						0.951	0.780	1.277	
MLP (ReLU)						0.979	0.822	1.280	
Support Vector Regression						0.989	0.828	1.286	
k-Nearest Neighbor Regression						0.876	0.915	1.322	
						Mean	0.879	0.840	1.283
						Min	0.649	0.780	1.256
						Median	0.914	0.825	1.278
						Max	0.989	0.915	1.322

Note. The cell shading indicates that the model achieves the best performance for the metric. Bold type indicates that the model outperforms all the reference models.

表5 Aggregation by the Parameters

	RMSE (Train)		RMSE (Val)		RMSE (Test)		Time (s)
	Mean	Min	Mean	Min	Mean	Min	Mean
Dimensions in Splines (d)							
- 2	0.724	0.631	0.801	0.747	1.281	1.223	121.078
- 3	0.733	0.633	0.813	0.753	1.286	1.226	135.471
- 4	0.723	0.624	0.801	0.752	1.279	1.226	149.866
- 5	0.715	0.643	0.790	0.757	1.268	1.228	163.955
- 1	0.719	0.596	0.798	0.755	1.277	1.229	105.676
A Number of Grids (K)							
- 50	0.707	0.596	0.791	0.749	1.266	1.223	152.772
- 10	0.721	0.650	0.797	0.752	1.275	1.226	121.719
- 20	0.740	0.631	0.814	0.747	1.294	1.226	129.998
Hidden Architectures							
- [3]	0.742	0.641	0.821	0.757	1.303	1.223	86.321
- [10]	0.722	0.643	0.797	0.757	1.272	1.226	95.459
- [10, 5]	0.722	0.623	0.801	0.747	1.280	1.226	129.469
- [50]	0.722	0.660	0.785	0.759	1.260	1.226	135.386
- [5]	0.712	0.633	0.793	0.758	1.270	1.229	87.191
- [5, 3]	0.754	0.624	0.841	0.749	1.313	1.229	121.999
- [20, 10, 5]	0.700	0.596	0.783	0.756	1.263	1.230	181.685
- [50, 20, 10]	0.711	0.650	0.782	0.752	1.261	1.232	243.442
A Number of Hidden Layers (L)							
- 1	0.724	0.633	0.799	0.757	1.277	1.223	101.040
- 2	0.733	0.623	0.814	0.747	1.291	1.226	126.988
- 3	0.705	0.596	0.782	0.752	1.262	1.230	211.243
Optimizer							
- Adamax	0.727	0.633	0.800	0.758	1.284	1.223	135.911
- AdamW	0.720	0.596	0.794	0.747	1.278	1.226	136.347
- Adam	0.714	0.623	0.789	0.749	1.271	1.226	137.518
- SGD	0.732	0.677	0.821	0.773	1.279	1.257	128.894
Scheduler							
- None	0.727	0.633	0.800	0.758	1.284	1.223	135.911
- Linear	0.720	0.596	0.794	0.747	1.278	1.226	136.347
- Exponential	0.714	0.623	0.789	0.749	1.271	1.226	137.518
Normalization							
- None	0.864	0.643	0.898	0.765	1.348	1.223	132.496
- Min–Max	0.691	0.623	0.782	0.762	1.263	1.232	133.139
- Percentile	0.691	0.596	0.778	0.752	1.263	1.245	135.541
- Clipping Norm.	0.692	0.624	0.778	0.747	1.261	1.245	135.727
- Sigmoid	0.702	0.676	0.784	0.764	1.268	1.258	136.877