

不確実な顧客需要の予測を目的とした混合ガンマ分布を仮定した深層分位点回帰

Deep Quantile Regression with Gamma Mixture Models for Forecasting Uncertain Customer Demand

北嶋 謙一^{†*} 今井 貴史[†]
Kenichi Kitajima Takashi Imai

1. 序論

近年、サプライチェーンの効率化や生産計画・在庫の最適化が求められる中で、それらを支える需要予測を、データに基づいて行う仕組みの重要性が一層高まっている。自動車や農業機械の部品業界においても同様であり、精度の高い需要予測は、製品の製造・供給の安定性とコスト削減を両立させるために、精度の高い需要予測が不可欠である。自動車部品や農業機械部品は、国内外に広く流通しており、その需要は、政治・経済の国際情勢に加えて、完成品メーカーやその上流に位置する加工メーカー、組立メーカーなど、サプライチェーン全体の経営戦略や生産計画など多岐に渡る要因の影響を受けて日々変動する。

しかし、実際の現場では、過去のデータを体系的に活用して予測を行う仕組みが十分に整備されておらず、多くの場合、担当者の経験や勘に基づいた属人的な判断が予測の主な手段となっている。顧客から定期的に提供される発注予定も、実際には急な変更やキャンセルが頻発するため、発注予定の信頼性の判断は担当者に委ねられている。このような担当者による属人的な予測は、担当者間での予測精度に一貫性がないだけでなく、担当者変更となった際は予測の再現性にも欠けるため、将来の突発的な需要変化に永続的に対応しきれないリスクをはらんでいる。

加えて、こうした部品需要の変動には「不確実性」が伴うため、単に点予測に基づいた従来手法では、突発的な発注増減への対応は困難である。こうした課題に対し、需要の不確実性そのものを定量的に捉える手法として、分位点回帰などのアプローチが注目されている。特に、近年では、複雑な需要のパターンを高精度に捉えられる深層ニューラルネットワークに用いた深層分位点回帰モデルも提案されている。例えば、Shenら(2021)は、ピンボールロス関数を用いた深層分位点回帰モデルについて解析を行っている。

本研究では、こうした背景を踏まえ、需要の不確実性を柔軟かつ高精度に捉えることを目的として、混合ガンマ分布に基づく深層分位点回帰モデルを提案する。これにより、需要分布全体を的確に把握するとともに、未観測要因が多い状況下においても、変動の大きい需要への対応力を向上させることを目指す。部品需要に見られる短期的傾向を効率的に捉えるため、モデルの基本構造として畳み込みニューラルネットワーク (Convolutional Neural Network; CNN) を導入する。また、ある時点の部品需要は前日や前週の同曜日からの影響があると考えられることを踏まえた二次元畳み込み層を組み込むことで、効率よく前日及び前週の同曜日との関連性を捉え、モデルの柔軟性と実用性を一層高

める。加えて、本モデルの学習過程における学習の安定性を実現するための損失関数を導入する。

本論文は以下のように構成されている。第 2 節では関連研究を整理し、本研究の位置づけを明らかにする。第 3 節では、ピンボールロス関数を用いる既存手法の問題点について述べる。第 4 節では、本研究における提案手法について詳述する。第 5 節では、実データを用いて検証を通じて、本手法の有効性を実証する。最後に、第 6 節で本研究のまとめと今後の課題について述べる。

2. 関連研究

本研究の位置づけを明確にするため、古典的な時系列手法、分位点回帰による不確実性の定量化、深層学習を用いたアプローチ、そして深層学習における学習停滞に関する関連研究を概観する。

需要予測は従来、Holt-Winters 法や ARIMA モデルといった統計的時系列モデルにより実施されてきた。これらはトレンドや季節性の抽出に優れ、Kumarら(2024)の事例にあるように製造業や小売業における実務上の課題に広く活用されている。しかし、突発的な変動や非線形関係への対応には限界がある。また、これらは基本的に点予測を目的としており、不確実性の定量的扱いには向いていない。この点を補完する手法として分位点回帰が提案されており、需要の不確実性を特定の分位点で直接予測することが可能である。Jha and Suresh(2019)は、スーパーマーケットの販売データに対し分位点回帰を適用し、特に上方リスクへの対応に有効であることを示した。ただし、分位点ごとにモデルの学習をする必要があるといった課題も指摘されている。

近年では、深層学習を用いた予測手法が注目されている。1次元 CNN を用いた分析の事例として、Namdariら(2023)による都市水の需要予測や、Langら(2019)による電力負荷の予測、Harbola and Coors(2020)による風速と風向きの予測などが挙げられる。一方で、これらの多くは点予測に留まり、予測の不確実性には対応していない。一方、Federico(2021)は複数の分位点を同時に出力する深層学習モデルを提案し、時系列依存性と不確実性の両立を試みている。また、Wuら(2023)の TimesNet では、時系列データを 2次元テンソルへと変換し、マルチスケールな周期性を捉えることで、多様な時系列タスクに対して高精度な予測を実現している。TimesNet は、周期性の抽出に高速フーリエ変換を用いており、周期をデータ駆動的に自動で決定している。一方、本研究で扱う部品の需要予測では、発注が行われる月曜日から土曜日の 6 日間という明確な周期が存在するため、周期の自動抽出は必須ではない。

ニューラルネットワークで分位点回帰を実現するための自然な方法として確率分布のパラメータを出力とするアプローチが考えられる。しかし、Seitzerら(2022)が指摘したように、対数尤度を用いた確率的回帰モデルにおいて、分

[†] 滋賀大学 Shiga University.

^{*} ゴーシューホールディングス Gohsyu Holdings, Inc.

散が大きいデータ点が学習にしばしば寄与しにくくなり、局所的かつ不適切な解に収束することがある。これに対し、彼らは分散に対する勾配の依存度を制御可能な β -NLL 損失関数を提案し、性能の改善と学習安定性の向上を実現している。ただし、この損失関数は正規分布を前提としており、ガンマ分布などの他の確率分布では機能しない。正規分布以外の分布を仮定した研究として、Sensoy ら(2018) による Evidential Deep Learning (EDL) が挙げられる。同手法は、従来の深層学習モデルが予測の確信度に対する不確実性を適切に定量化できないという課題に着目している。この問題に対処するため、EDL では、予測がディリクレ分布に従うことを仮定して、(1)式に定義されるような期待二乗誤差に基づく損失関数を導入し、予測誤差項と分散項の双方を最小化するよう設計されている。

$$\mathbb{E}[p_{ij}^2] = \mathbb{E}[p_{ij}]^2 + \text{Var}(p_{ij}) \quad (1)$$

このように、需要予測とその不確実性を扱う研究や、それに付随する学習安定性に関する研究などは多岐にわたる。本研究ではそれらを参考に、不確実性を考慮した需要予測を、学習安定性を担保しつつ実現する。

3. 既存手法の問題点

本節では従来の分位点回帰で中核となるピンボールロス損失関数に焦点を当て、定義とその特性について概観するとともに、需要予測モデルの整合性に関する課題を明らかにする。

3.1 分位点回帰の基本

分位点回帰とは、従来の最小二乗法と異なり、目的変数の条件付き分位点を直接推定する回帰手法である。分位点回帰では、観測値 y に対して、特定の分位点 $\tau (0 < \tau < 1)$ を予測する。 τ 分位点 $Q_\tau(x)$ は、 $\mathbf{P}(Y_t \leq Q_\tau(x) | \mathbf{X} = \mathbf{x}) = \tau$ を満たすように定義される。不確実性を推定できるため、在庫管理や医療、金融などのリスクを伴う領域において有効とされる。

3.2 ピンボールロス関数の定義と性質

分位点回帰において広く用いられている損失関数は、ピンボールロス関数である Koenker(2005)。 τ 分位点を求める場合のピンボールロス関数は、 $\tau \in (0,1)$ に対して(2)式のように定義される。

$$\begin{cases} \tau(y_i - \hat{y}_i) & \text{if } y_i \geq \hat{y}_i \\ (1 - \tau)(\hat{y}_i - y_i) & \text{if } y_i < \hat{y}_i \end{cases} \quad (2)$$

ここで y_i は実測値、 \hat{y}_i は予測値を示す。この損失関数は、目標とする分位点に応じて過小予測と過大予測のペナルティを異なる重みで与えることにより、特定の分位点を学習する仕組みとなっている。例えば、 $\tau = 0.9$ の場合、予測が実測より小さい場合には大きな損失が課される一方で、過大予測には比較的小さなペナルティが与えられる。なお、 $\tau = 0.5$ の場合、ピンボールロス関数は平均絶対誤差 (Mean Absolute Error; MAE) と同じ形となり、MAE は過大評価と過小評価の誤差に対して対称的にペナルティを与える。MAE は外れ値や大きなノイズの影響を受けにくい (ロバスト) 性質を持つ損失関数として知られていることから、ピンボールロスもまた、 $\tau = 0.5$ においては同様の性質を有する。ただし MAE と異なり、ピンボールロスは τ を

変更することで、任意の分位点を学習することができ、区間予測への応用が可能である。また、ピンボールロスは原点付近で不連続であるものの、凸関数であるため、最適化問題としても扱いやすく、勾配ベースの手法によって効率的に学習が可能である。

3.3 モデルとの整合性とその問題点

Tagasovska and Lopez-paz(2019) が述べているように、ピンボールロス関数を用いた分位点回帰は目的変数に分布を仮定せず、特定の条件付き分位点を直接学習する手法であり、分布の非対称性や外れ値の存在下でも安定した推定が可能である。その一方で、分布を明示的に仮定せず、損失関数が真の分布をどのように反映しているかを明示できないことにより、いくつかの問題点も存在する。

第 1 に、Tagasovska and Lopez-paz(2019) も指摘しているように、異なる分位点予測間の整合性が保証されないという問題がある。モデルが出力する分位点は確率分布に基づく一貫性を持たないため、例えば 90% 分位点より 95% 分位点が小さくなるといった理論的に矛盾する予測が発生し、解釈に困難が生じる可能性がある。

第 2 に、条件付き分布全体をモデル化できないため、ある時点の需要のばらつきや分散構造を明示的に表現することが困難である。これは、安全在庫の設計や在庫戦略の策定など、不確実性を前提とした意思決定を行う場面において大きな制約となる。

第 3 に、分位点ごとに個別にモデルを学習する場合、学習コストが増大するだけでなく、分位点間での重みの解釈性や出力結果の一貫性が失われやすい点である。製品ごとに異なる分位点が要求されるような実務環境では、モデルの再利用性や拡張性の面でも不利に働く可能性がある。

これらの問題点は、分布を出力とすることで解消される。加えて、分布を出力とするモデル設計は、将来的な研究の発展にも柔軟に対応できる。例えば、今回構築したモデルを予測タスクだけではなく各製品の需要パターンに基づいたクラスタリング等に適用することも可能となる。

これらの理由から、需要予測の不確実性が高い環境下で実務的にモデルを使用する場合、分布仮定に基づき分布全体を一貫して予測可能なモデルの構築が有望である。

4. 提案手法

本節では、第 3 節で明らかにした課題を踏まえた提案手法を紹介する。

4.1 混合ガンマ分布を仮定したモデル設計

第 3 節で述べたように、ピンボールロスを用いた分位点回帰は確率分布を前提としない点で、柔軟性がある一方、モデル出力の整合性や解釈性に課題が残る可能性がある。そこで本研究では、これらの課題を克服するため、目的変数が従う分布を明示的に仮定したモデルを設計する。需要は非負の連続値であるため、目的変数が従う分布の候補としてはガンマ分布が考えられる。しかし、ガンマ分布を用いた予備実験において、上限分位点が過剰に大きくとられる結果が確認された。これは、実データにおける多峰性や局所的な需要変動を捉えるには単一のガンマ分布では表現力が不足しており、需要の不確実性を十分に表現することができなかったためと考えられる。そこで本研究では目的

変数 y が従う分布として、混合ガンマ分布を仮定する。混合ガンマ分布の確率密度関数は以下の(3)の式で与えられる。

$$p(y|\eta_{1:m}, \alpha_{1:m}, \beta_{1:m}) = \sum_{m=1}^M \eta_m \text{Gamma}(y|\alpha_m, \beta_m) \quad (3)$$

ここで、 M は混合要素の個数、 η_m は第 m 成分の混合比率、Gamma はガンマ分布の確率密度関数、 α_m および β_m はそれぞれ第 m 成分のガンマ分布における形状パラメータおよび比率パラメータを表す。混合ガンマ分布は、複数の異なるガンマ分布の加重平均によって構成され、分布の多峰性を柔軟に表現することが可能である。本研究における提案モデルでは、ニューラルネットワークにより混合比率 $\eta_{1:m}$ 、各成分の形状パラメータ $\alpha_{1:m}$ 、および比率パラメータ $\beta_{1:m}$ を同時に出力する構成とし、これにより確率密度関数全体を出力値としてモデル化することが可能な予測機構を実現する。

4.2 損失関数の設計

混合ガンマ分布を出力とする場合の自然な損失関数としては、負の対数尤度 (Negative Log-Likelihood; NLL) が考えられる。NLL は以下の(4)で表される。

$$\text{NLL} = -\log p(y_{\text{true}} | \eta_{1:m}, \alpha_{1:m}, \beta_{1:m}) \quad (4)$$

ここで、 y_{true} は観測された発注量である。しかし、NLL をそのまま適用した場合、モデルは観測値との整合性よりも、分布の分散を大きくとることで損失を減らそうとする傾向が見られた。これは、Seitzer ら (2022) も指摘しているように、学習過程において極端に裾の広い分布を生成することで、真の分布からの乖離を招き、良好な解への収束を妨げていることを意味する。そこで本研究では、このような挙動を抑制するため、NLL と併用して、分布の分散の大きさにペナルティを与える正則化効果を持つ期待二乗誤差 (Expected Squared Error; ESE) 損失関数を導入した。ESE は以下の(5)式で定義される。

$$\begin{aligned} \text{ESE} &= \mathbb{E}[(y - y_{\text{true}})^2] \\ &= \mathbb{E}[(y - \mu) + (\mu - y_{\text{true}})]^2 \\ &= (y_{\text{true}} - \mu)^2 + \sigma^2 \end{aligned} \quad (5)$$

ここで μ と σ^2 は(6)式に示すように、それぞれ出力された混合ガンマ分布の平均と分散を表す。

$$\mu = \sum_{m=1}^M \eta_m \frac{\alpha_m}{\beta_m}, \quad \sigma^2 = \sum_{m=1}^M \eta_m \frac{\alpha_m}{\beta_m^2} \quad (6)$$

これにより、予測の精度と不確実性のバランスを同時に考慮した学習が可能となる。需要予測における分布整合性と学習の安定性を同時に考慮するために、本研究ではNLLとESEを統合したハイブリッド損失関数を導入する。ハイブリッド損失関数は以下の(7)式によって定義される。

$$L_{\text{hybrid}} = (1 - \lambda) \times L_{\text{NLL}} + \lambda \times L_{\text{ESE}} \quad (7)$$

ここで、 L_{NLL} は NLL 損失、 L_{ESE} は ESE 損失である。 $\lambda \in [0, 1]$ は両損失の相対的な寄与を制御する重み係数であり、モデルの学習目的によって調整可能である。NLL によって予測分布全体の整合性を最大化する一方、ESE は分散が過剰に拡張されることを抑制し、予測区間の幅と中心の両方に対してバランスの取れた学習が可能となる。このような

ハイブリッド構造により、実践上求められる「分布的な正しさ」と「予測区間の実用性」のトレードオフを制御可能とし、安定かつ整合性のある予測性能を実現する。

4.3 学習プロセスと重み調整戦略

本研究では、提案モデルの学習過程において、NLL のみによる最適化では分散を過度に広くとることで損失を減らそうと傾向があることを考慮し、ESE も導入したハイブリッド損失関数を使用して、段階的な損失関数の重み調整による 2 段階の学習プロセスを採用する。具体的には以下の 2 段階からなる。

(1)事前学習 (Pretraining)

ハイブリッド損失関数の重み λ を非ゼロの値 (本研究では $\lambda = 0.5$) に設定し、NLL と ESE の双方が同程度に最適化されるように学習を行う。この設定により、分布整合性と学習の安定性とのバランスを保ち、妥当な初期パラメータの獲得を目指す。これにより、モデルは分散を過度に広くとった解に陥ることなく、学習初期の探索範囲を広げることが可能となる。

(2)再学習 (Fine-tuning)

次の段階では、ハイブリッド損失関数の重みを $\lambda = 0$ に設定し、NLL のみに基づいて学習を進める。これにより、事前学習で得られた初期パラメータを基に、より正確に目的変数の確率密度関数を近似できるよう学習を行う。この再学習により、分布整合性の強化と予測精度の向上が期待される。

この 2 段階の学習戦略により、異なる性質を持つ損失関数を補完的に活用し、最適化過程における安定性と性能の両立を実現する。

4.4 カレンダー構造を適用した層設計

部品需要には曜日や月、祝日などのカレンダー要因に起因する周期的な変動がしばしば観察される。これらのパターンは単純な時系列変動にとどまらず、曜日や週単位で繰り返される需要リズムとしても表れると考えられるため、深層学習モデルにおいてこうしたカレンダー構造を適切にモデル化することは予測精度の向上に資する。CNN を用いた多くの従来の時系列予測手法では 1 次元畳み込み (Conv1D) が用いられており、曜日や週単位の需要リズムを捉えるためには、カーネルサイズを非常に大きくする必要があり、本研究では、こうした構造的特徴を効率的に捉えるために、カレンダー構造を明示的に反映した二次元畳み込み層を設計し、曜日軸と週軸の両方における需要パターンを同時に抽出することを可能にする。

図 1 に示すように、本層の入力は 1 次元形状 $B \times (W \cdot D) \times F$ である。 B はバッチサイズ、 W は週数 (本研究では 1 か月相当の過去データに基づき予測を行うため $W = 5$)、 D は曜日数 (本研究では月曜日から土曜日を使用するので、 $D = 6$)、 F は特徴量数である。これを二次元形状 $B \times W \times D \times F$ に変換し、畳み込みを行う。

さらに、時間的な局所依存性をより忠実にモデル化するために、入力系列に対するパディング処理にも工夫を加える。通常、畳み込み操作の際にはゼロや固定値を用いたパ

3/1	3/2	3/3	...	4/1	4/2	4/3
-----	-----	-----	-----	-----	-----	-----

	Pad	Mon	Tue	Wed	Thu	Fri	Sat
Pad	0	0	0	0	0	0	0
Week1	0	3/1	3/2	3/3	3/4	3/5	3/6
Week2	3/6	3/8	3/9	3/10	3/11	3/12	3/13
Week3	3/13	3/15	3/16	3/17	3/18	3/19	3/20
Week4	3/20	3/22	3/23	3/24	3/25	3/26	3/27
Week5	3/27	3/29	3/30	3/31	4/1	4/2	4/3

赤枠はカーネル
黄色はパディング

図 1 CalendarConv2D の概要

ディング処理がされるが、本研究では、自然な時間的連続性を維持するためカレンダー構造を反映したパディングを採用している。

この層を CalendarConv2D と呼ぶこととする。このように、本研究で提案する CalendarConv2D は、従来の Conv1D に比べて、カレンダー的な周期構造を効率的にモデル化することを可能とし、周期的関係の学習効率と予測性能の両面において優位性を持つと考えられる。

5. 実験

本節では、提案手法の有効性を検証するため、実データを用いた予測実験を行う。ベースライン手法にピンボールロスペースモデルを用いる。本実験の目的は、提案手法が実際の製品の需要予測において有効に機能するかを評価し、実務での運用可能性を検証することである。

5.1.1 モデル構造

図 2 に、本研究で用いるモデル構造を示す。本モデルは、日次・週次・月次の 3 系列の入力を段階的に統合する多入力型の深層ネットワークで構成されている。日次系列の学習ブロックの出力を x_1 、週次系列の学習ブロックの出力を x_2 である。

まず日次ブロックでは、日次系列データと、説明変数が休日かどうかを識別するための休日情報を入力とし、CalendarConv2D を適用する。この層は、曜日構造および週単位の周期性を考慮した特徴抽出を目的としており、曜日や週間に見られる特有の需要パターンを反映できるように設計されている。カーネルサイズには 2×2 を採用することで、前日および 1 週間前の同曜日に対する時系列的な局所依存性を同時に捉えることが可能となる。これにより、需要の短期的な変動と周期的傾向を効果的に学習することが可能である。週次ブロックでは x_1 と週次系列データ及び日次系列データを Conv1D 層に通して週単位での時系列の変化を捉える。日次ブロックと週次ブロックではスキップ接続を導入することで、学習の安定性を高めている。最後に月次ブロックでは x_2 に加えて月次系列データ、日次系列データ、予測日が休日かどうかを予測するための休日データを入力し、全結合層で統合的に処理する。月次ブロックから出力された特徴量は、形状パラメータ $\alpha_{1:m}$ 、比率パラメ

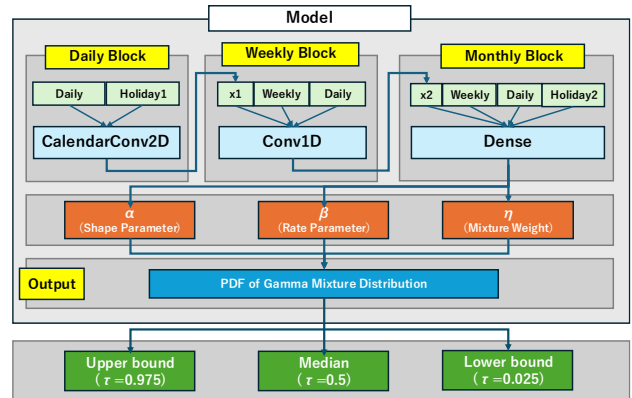


図 2 モデル構造

ータ $\beta_{1:m}$ 、および混合比率 $\eta_{1:m}$ に変換され、混合ガンマ分布の確率密度関数を最終出力とする。

出力された確率密度関数に基づき、必要な分位点を計算する。本研究では、予測区間の設定として、下限分位点・上限分位点をそれぞれ 2.5% と 97.5% とし、95% 予測区間を対象とした評価を行う。これは実運用において、生産計画や在庫の最適化には概ね 95% 程度の予測精度の信頼水準が求められると判断したことによる。また、中央値 (50% 分位点) も求めることで、予測値が実測値の傾向を捉えられているかを確認する。

5.1.2 評価指標

提案手法によって出力される分位点予測の性能を定量的に評価するため、以下の 2 種類の評価指標を用いる。

① 被覆率 (Coverage Rate)

被覆率は、指定された予測区間 $[L, U]$ に観測値 y が含まれている割合を示す。 n 個の予測に対して、次の (8) 式で定義される。

$$\text{Coverage Rate} = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(y_i \in [L_i, U_i]) \times 100 \quad (8)$$

ここで、 y_i は i 番目の実測値、 $\mathbf{1}(\cdot)$ は指示関数、 L_i および U_i はそれぞれ下限および上限分位点である。被覆率は理論上の信頼水準にどの程度一致しているかを示す簡潔な指標であり、予測区間の過小・過大評価の診断に有効である。値が大きいかほど高い精度と解釈する。ただし、極端に広い予測区間を出力した場合でも被覆率は高くなるため、実用性を適切に反映しない可能性がある。

② Winkler スコア

Winkler スコアは、予測区間の幅と、観測値が区間外にある場合の外れ度合いの両方を考慮する指標であり、以下の (9) 式のように定義される (Winkler, 1972)(Gneiting and Raftery, 2007)。

$$\begin{aligned} \text{Winkler Score}(L, U; y) \\ = (U - L) + \frac{2}{\alpha} (L - y) \mathbf{1}\{y < L\} + \frac{2}{\alpha} (y - U) \mathbf{1}\{y > U\} \quad (9) \end{aligned}$$

ここで、 y は実測値、 L および U はそれぞれ予測区間の下限および上限、 α ($0 < \alpha < 1$) は予測区間の被覆率に対する許容誤差率を表す。本研究では $\alpha = 0.1$ とする。観測値が予測区間内にある場合には区間幅のみが評価対象となり、外れた場合にはその距離に応じてペナルティが加算される。

表 1 説明変数一覧

粒度	分類	変数例	変数数	内容/特徴
月次	経済指標	輸出価格指数	12	外部環境のマクロ的変動を反映
	顧客内示	製品Bの月次内示量	17	顧客企業からの月次見込みを反映
	顧客内示・実績差異	製品Bの内示実績差異	17	月次見込み数と月次実績数の差異を反映
週次	経済指標	ディーゼル価格	1	外部環境の週次での変動を反映
	発注データ	製品Bの週次発注量	28	週単位での発注トレンドを反映
日次	経済指標	為替、日経平均株価	3	外部環境のミクロ的変動を反映
	発注データ	製品Bの日次発注量	29	製品間の連動性を反映
	株価	企業Bの株価	11	関連企業間の株価の変動を反映
	カレンダー要因	祝日	2	目的変数と説明変数の日付が休日かを反映
	カレンダー要因	曜日、日付	9	営業日や周期性を反映

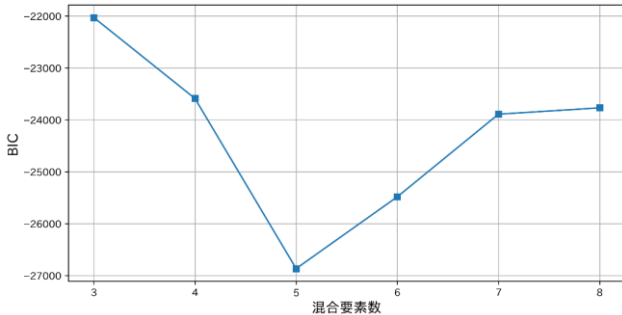


図 3 BIC に基づく最適な混合要素数 η の選定

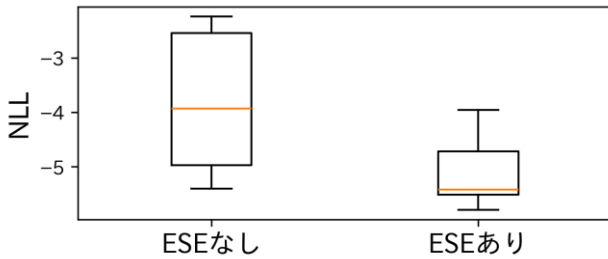


図 4 ESE 導入有無における学習終了後の対数尤度 (NLL)

値が小さいほど高精度と判断される。直観的な解釈はやや難しいものの、予測区間の妥当性を厳密に評価できる総合的な指標である。

従って、本研究では「被覆率が高く、かつ Winkler スコア が小さいほど、予測区間の性能が優れているモデルである」と解釈する。

5.1.3 使用データセット

本実験では、株式会社ゴーシューホールディングスより提供された農業機械業界向けの製品 A の日次の発注量データを用いる。今回使用したデータは 2020 年 1 月 6 日から 2024 年 10 月 31 日までの発注データである。訓練データには 2020 年 1 月 6 日から 2024 年 7 月 6 日までのデータを用いて、テストデータには 2024 年 7 月 8 日から 2024 年 10 月 31 日までのデータを用いる。使用する説明変数は、表 1 に示すとおりである。目的変数には、発注から生産完了までに要するリードタイムを考慮し、1 週間後の発注量を用いる。これは実務上の調達・生産計画において重要なタイムスパンに相当する。

表 2 ベースライン手法と提案手法における被覆率と Winkler スコア

評価指標	ピンボールロス	提案手法
被覆率(%)	90.5 (4.39)	92.50 (2.54)
Winkler スコア	7239.57 (449.05)	5292.03 (574.41)

※()は標準偏差

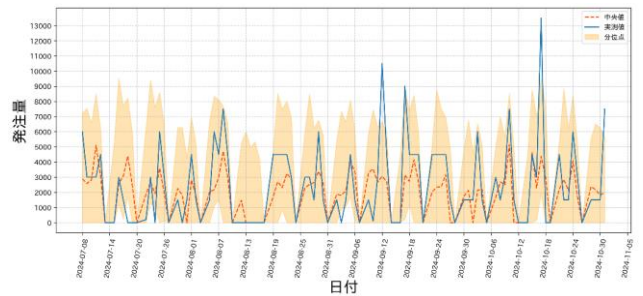
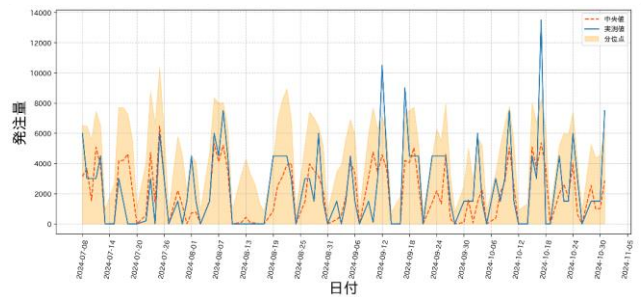


図 5 ベースライン手法 (上) と提案手法 (下) の予測結果

5.2 実験結果

5.2.1 混合要素数の選定

本研究においては、混合ガンマ分布の混合要素数 M を決定するために、 $M = 3$ から $M = 8$ の範囲でモデルを構築し、ベイズ情報量規準 (Bayesian Information Criterion; BIC) に基づいて評価を行った。BIC は以下の(10)式で定義される。

$$BIC = -2 \log(L) + k \log(n) \quad (10)$$

ここで、 k はパラメータの次元、 L は尤度、 n はデータ点の個数を表す。その結果、BIC が最小値を示した $M = 5$ を最適な混合要素数として採用した (図 3 参照)。

5.2.2 ESE 導入による学習安定化の検証

提案手法における ESE の導入が、分布の分散が過剰に大きくなる傾向を抑制するかを検証する。図 4 は、ESE を導入した場合および導入しない場合のそれぞれについて、同一条件下で 10 回実験を実施した際の損失値を可視化したものである。図が示すように、ESE を導入しなかった場合には、損失の中央値が相対的に高く、ばらつきも大きい結果となった。このことは、学習が初期値に大きく依存し、収束先にばらつきが生じる傾向があることを示唆している。一方で、ESE を導入した場合には、損失の中央値が小さく、

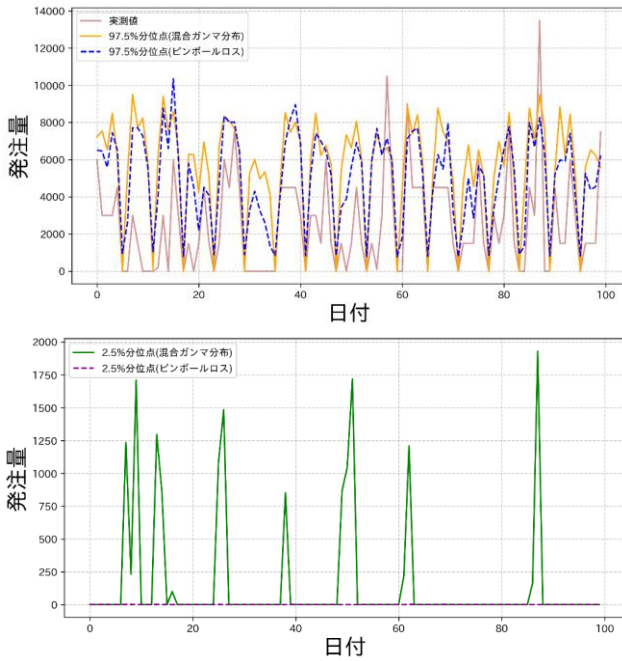


図 6 分位点 97.5%差異 (上) と分位点 2.5%差異 (下)

かつばらつきも抑えられている。これは、初期値に対する頑健性が向上し、学習結果の安定性および再現性が高まっていると考えられる。さらに、損失の低さは、ESE の導入によってより良好な解に近い解に到達している可能性を示唆している。以上の結果から、ESE の導入は良好な解に収束することを促し、学習の安定性を向上させる有効な手段であると考えられる。

5.2.3 予測精度の評価

図 5 にベースライン手法による予測結果と、提案手法による予測結果を示す。青実線が実測値、赤点線が中央値、オレンジ色の領域が 95% 予測区間である。併せて表 2 に、ベースライン手法と提案手法を同一条件下で 10 回実施した際の被覆率と Winkler スコアの平均値と標準偏差を示す。両手法ともに、極端な発注量を除けば、中央値および 97.5% 分位点は実測値の変動をおおむね適切に捉えていることが確認される。加えて、表 2 より、被覆率および Winkler スコアの両評価指標の平均値において、提案手法がベースライン手法を上回る結果を示した。この結果は、提案手法が予測区間の幅を不必要に広げることなく、適切な範囲で実測値を包含できていることを意味しており、分布整合的な予測構造と損失関数設計が予測精度に対して有効に寄与していることを示している。また、両指標の標準偏差について、被覆率は提案手法の方が小さいことから、精度が安定していると解釈できる。一方で、Winkler スコアの標準偏差については、両手法ともに同程度の結果が得られた。しかし、次段落でも述べるように、ベースライン手法では下限区間の多くの期間で予測値が 0 近傍に位置している。一方で、提案手法では特定の期間において下限分位点が非ゼロの値を取り、実際の需要のばらつきをより忠実に捉えた予測が行われている。これにより、Winkler スコアの標準偏差は両手法で同程度であっても、その内実は大きく異なることが示唆される。すなわち、提案手法は同等のスコア変動を維持しつつ、より正確に需要の不確実性

表 3 CalendarConv2D (カーネル 2)、Conv1D (カーネル 4)、Conv1D (カーネル 7) における被覆率と Winkler スコア

評価指標	CalendarConv2D	Conv1D (カーネル4)	Conv1D (カーネル7)
被覆率(%)	92.75 (2.41)	92.15 (2.31)	92.7 (4.14)
Winkler スコア	5304.67 (471.41)	5873.73 (630.25)	5701.87 (656.22)

※()は標準偏差

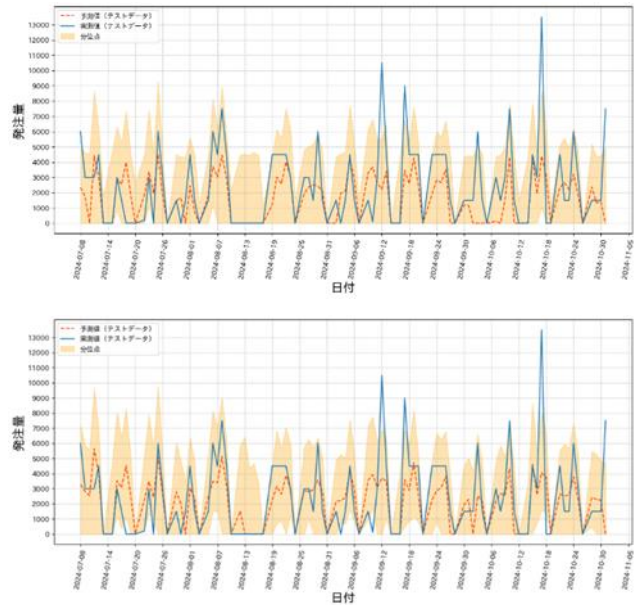


図 7 Conv1D (カーネル 4) (上) と Conv1D (カーネル 7) (下) の予測結果

を考慮した予測区間を提供していると解釈できる。このことは、単なるスコア値の比較だけでは捉えきれない、予測モデルの構造的な違いと実務的有用性を示唆している。

5.2.4 上下限区間の表現力の評価

図 6 に、それぞれ 97.5% 分位点および下限 2.5% 分位点の予測結果を示す。青点線がベースライン手法における 97.5% 分位点、オレンジ実線が提案手法における 97.5% 分位点、茶色の薄線が実測値である。紫点線がベースライン手法における 2.5% 分位点、緑実線が提案手法における 2.5% 分位点である。97.5% 分位点に関しては両手法ともに類似した予測傾向を示しており、実測値が示す傾向を概ね適切に捉えている。これは提案モデルが、分布仮定をしないピンボールロスベースモデルと少なくとも同程度の表現力を有することを示している。一方、下限分位点については、両手法の間で顕著な差異が認められる。ベースライン手法では、ほぼ全ての期間において予測値が 0 近傍に位置しており、下方リスクを適切に予測できていない。一方、提案手法は、特定の期間において下限分位点が非ゼロの値を取り、需要のばらつきに応じた予測がなされていることが分かる。これは、混合ガンマ分布に基づく分布整合的な予測が、より柔軟に下方の不確実性を捉えていることを示している。

5.2.5 CalendarConv2D と Conv1D の比較

CalendarConv2D の有効性について検証を行うために、混合ガンマ分布を仮定した提案手法において、CalendarConv2D を導入せず、代わりに Conv1D 層のみを用いて予測を行った場合の予測を行う。比較対象となる Conv1D については、カーネルサイズを 4 および 7 の 2 通りで設定している。カーネルサイズ 4 は、CalendarConv2D において使用した 2×2 の畳み込みと同じ日数を Conv1D が参照するように設定したものである。一方、カーネルサイズ 7 は、本研究において 1 週間を 6 営業日と定義しているカレンダー構造を踏まえ、前日および 1 週間前の同曜日を 1 次元上で同時に捉えることができるよう設計したものである。このように設計された Conv1D との比較を通じて、CalendarConv2D が周期的な時間構造やカレンダー要因をより効果的に学習・反映できているかを評価する。図 7 に、Conv1D における予測結果を示す。表 3 には、CalendarConv2D を導入した場合と、Conv1D のみを用いた場合に、同一条件下で 20 回実施した際の被覆率および Winkler スコアの比較結果を示す。結果として、CalendarConv2D といずれの Conv1D 構造との比較において、被覆率の平均値と標準偏差にほぼ同等の結果が得られた。一方で、Winkler スコアについては、CalendarConv2D がいずれのカーネルサイズの Conv1D よりも平均値と標準偏差の両方で有意に優れた結果を示した。このことは、予測区間に関しては、CalendarConv2D は Conv1D と同程度の予測能力を有すると考えられる。その一方で、予測区間の幅や外れ度合いを含む不確実性を詳細に捉える点においては、CalendarConv2D の方が初期値に依らず高い予測能力を有することを示唆している。加えて図 7 より、下限分位点の予測挙動に着目すると、カーネルサイズ 4 の Conv1D では、予測された下限分位点が、多くの期間において 0 近傍に位置する傾向が確認された。一方で、カーネルサイズ 7 の Conv1D および CalendarConv2D では、下限分位点が需要のばらつきに応じて適切に変動していることが確認された。この結果は、カーネルサイズ 4 の Conv1D においては、前週同曜日の情報が特徴量に含まれておらず、モデルが保守的な推定を行い、需要の下方リスクを十分に予測できなかったことが示唆される。一方で、カーネルサイズ 7 の Conv1D と CalendarConv2D はいずれも曜日と週の構造的要因を同時に捉える情報を入力として保持しており、下方リスクを含めた周期的な需要パターンをより包括的に学習することが可能となっていたと考えられる。以上より、少なくとも CalendarConv2D は、予測区間の中心傾向に加え、下限分位点における不確実性の予測において、大きなカーネルサイズの Conv1D と同程度以上の予測能力を有していると考えられる。

6. 結論

本論文では、不確実な顧客需要の予測を目的として、株式会社ゴーシューホールディングス提供のデータを用いた需要予測モデルの構築を行った。需要予測における不確実性の定量化を目的として、CNN をベースとした深層分位点回帰モデルを構築した。需要に影響を与えらるる曜日や週単位の周期性構造を効率的に学習可能とするために、カレンダー構造を明示的に組み込んだ二次元畳み込み層である CalendarConv2D を設計した。さらにモデルの出力

に対しては、混合ガンマ分布を仮定することで、予測値の不確実性を確率的に表現した。また、学習の安定性を実現するために、分布の整合性と分散に対する正則化の両方の機能を有するハイブリッド損失関数を導入し、予測区間の信頼性とモデルの安定性を両立する設計とした。

ピンボールロスを使用した深層分位点回帰モデルをベースライン手法として、提案手法の性能を評価した結果、予測区間はより適切に実測値を包含しており、分布整合的かつ安定した学習を行うことができることを確認した。加えて、Conv1D との精度比較を行った結果、CalendarConv2D は大きなカーネルサイズの Conv1D と同程度以上の不確実性予測能力を有することが確認された。

しかしながら、提案手法の被覆率が、本研究で設定した 95% 予測区間を下回る結果が得られたことについては、引き続きモデルの表現力向上や突発需要を説明できるような変数の導入を検討する必要があると考えられる。

分布を出力としたモデル設計を行うことで、統計モデルに基づくデータマイニング手法と組み合わせることが可能となる。例えば、発注傾向が類似する製品のグルーピングを目的としたモデルベースクラスタリングへの応用が考えられる。需要パターンに基づいてクラスタリングを行うため、本研究で構築した混合ガンマ分布 CNN モデルを混合成分とする混合分布モデルを構築し、クラスタリング対象の製品すべてに対してそのモデルを当てはめる。そして、各製品について最も混合割合が大きいモデルをその製品が属するクラスターと見なすことでクラスタリングを行うことができる。このような応用は、最適在庫の設定や、生産計画立案における意思決定を支援する知見を見出すことに有用である。

参考文献

- [1] Guahao Shen, Yuling Jiao, Yuanyuan Lin, Joel L. Horowitz, Jian Huang, "Deep quantile regression: mitigating the curse of dimensionality through composition", arXiv:2107.04907 (2021).
- [2] Lalji Kumar, Sudhakar Khedlekar, U. K. Khedlekar, "A comparative assessment of holt winter exponential smoothing and autoregressive integrated moving average for inventory optimization in supply chains", Supply Chain Analytics, 8 (2024).
- [3] Swati Jha, Naryanarao Suresh, "Forecasting demand at supermarkets using quantile regression", International Journal of Mechanical Engineering and Technology, pp.59-71 (2019).
- [4] Hossein Namdari, Ali Haghghi, Seyed Mohammad Ashrafi, "Short-term urban water demand forecasting: application of 1D convolutional neural network (1D CNN) in comparison with different deep learning schemes", Stochastic Environmental Research and Risk Assessment, (2023).
- [5] Christian Lang, Florian Steinborn, Oliver Steffens, Elmar W. Lang, "Electricity load forecasting - An evaluation of simple 1D-CNN network structures", (2019).
- [6] Shubhi Harbola, Volker Coors, "One dimensional convolutional neural network architecture for wind prediction", Energy Conversion and Management 198, 111824 (2019).
- [7] Federico Garza Ramirez, "Neural Multi-quantile forecasting for optimal inventory management", (2021).
- [8] H. Wu, T. Hu, Y. Liu, H. Zhou, J. Wang, M. Long, "Timesnet: temporal 2D-variation modeling for general time series analysis", Proceedings of the International Conference on Learning Representations (ICLR), (2023).
- [9] Murat Sensoy, Lance Kaplan, Melih Kandemir, "Evidential deep learning to quantify classification uncertainty", Proceedings of the 32nd Conference on Neural Information Processing Systems (NeurIPS), (2018).
- [10] M. Seitzer, A. Tavakoli, D. Antić, G. Martius, "On the pitfalls of heteroscedastic uncertainty estimation with probabilistic

- neural networks,” Proceedings of the International Conference on Learning Representations (ICLR), (2022).
- [11] R. Koenker, “Quantile Regression”, Cambridge University Press, (2005).
- [12] Natasa Tagasovska, David Lopez-Paz, “Single-Model Uncertainties for Deep Learning”, Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS), (2019).
- [13] Winkler, R. L, “A decision-theoretic approach to interval estimation”, Journal of the American Statistical Association, 67(337), 187–191, (1972).
- [14] Gneiting, T., & Raftery, A. E., “Strictly Proper Scoring Rules, Prediction, and Estimation”, Journal of the American Statistical Association, 102(477), 359–378, (2007).