

死亡率の長期トレンドリスク評価のための異常値耐性のあるニューラルネットワークモデル

明治大学大学院 先端数理科学研究科 現象数理学専攻
博士前期課程2年 富山 蓮

【研究の背景と概要】

- 死亡率を予測する際に用いられる代表的なモデルがLee-Carter(LC)モデルであり、年齢 x 、歴年 t の死亡率 $m_{x,t}$ を次のように表すモデルであるが、長寿リスク評価や保険負債の経済価値評価に必要な死亡率の長期トレンドリスク評価においては、データの非線形性表現能力と共に、post-COVID-19で重要性が増す異常値耐性を持つモデルが求められる。

$$\log(m_{x,t}) = \alpha_x + \beta_x \kappa_t + \varepsilon_{x,t}, \varepsilon_{x,t} \sim N(0, \sigma^2)$$

(α_x : 対数死亡率平均, β_x : κ_t に対応する年齢別感応度, κ_t : 死亡率改善度曲線, $\varepsilon_{x,t}$: 平均0のノイズ項)

- 本研究では、再構成型の異常検知のためのニューラルネットワーク(NN)の一種であるLong Short-Term Memory Autoencoder(LSTM-AE)を利用したLCモデルの非線形拡張モデルを提案する。
- LCモデルは特異値分解(SVD)による次元削減の第一段階と、得られた時系列成分への時系列モデルの当てはめの二段階で推定されており、年齢成分と時系列成分パラメータの不整合性が懸念される。本研究では、LSTM-AEを用いて、パラメータ間の整合性を保つ一段階推定が可能で、リスク管理に不可欠な解釈可能性を持つ死亡率推定モデルを提案する。このモデルは、優れた初期値頑健性を持ち、死亡率曲線の自然な滑らかさを保った長期外挿を実現する。

提案モデル(LSTM-AEを利用した非線形拡張モデル)

■ LSTM-AEを利用したLCモデルの非線形拡張(再構成)

本研究は、Mirza & Cosan(2018)によって提案されたLSTM-AEを利用した異常検知モデルを利用しており、DecoderモジュールにおけるLSTMへの入力ベクトルがEncoderで得られた特徴量ベクトルで固定される。

$t = 1, \dots, d$ に対して、入力ベクトル $\mathbf{u}_t := (\log m_{1,t}, \dots, \log m_{X,t})$ と定義すると、Encoder関数: f^{enc} を利用して、特徴量ベクトル $\mathbf{r}$ は次のように計算される。

$$\mathbf{r} = f^{enc}(\mathbf{u}_1, \dots, \mathbf{u}_d)$$

DecoderモジュールにおけるLSTMの出力を $\mathbf{h}_t$ とすると、Decoder関数: f^{dec} を利用して、最終的な出力 $\hat{\mathbf{u}}_t$ は得られた特徴量ベクトル $\mathbf{r}$ より次のように表現することができる。

$$\hat{\mathbf{u}}_t = \log(\widehat{\mathbf{m}}_t) = f^{dec}(\mathbf{h}_{t-1}, \mathbf{r})$$

■ LSTM-AEを利用したLCモデルの非線形拡張(グラフ構造)

図1が提案モデルになり、最終的な出力が年齢成分と時系列成分を非分離とし双線形性を持たない構造にするためLSTMの出力と最終的な出力の間に層を一つ重ねている。

■ LSTM-AEを利用したLCモデルの非線形拡張(予測)

再構成によって得ることができたLSTM-AEの出力 $\mathbf{h}_t$ を外挿することで将来予測を行う。

DecoderモジュールにおけるLSTMの出力を表す関数を f_{LSTM}^{dec} とすると $\mathbf{h}_t$ は次のような再帰構造を表現できる。

$$\mathbf{h}_t = f_{LSTM}^{dec}(\mathbf{h}_{t-1}, \mathbf{r})$$

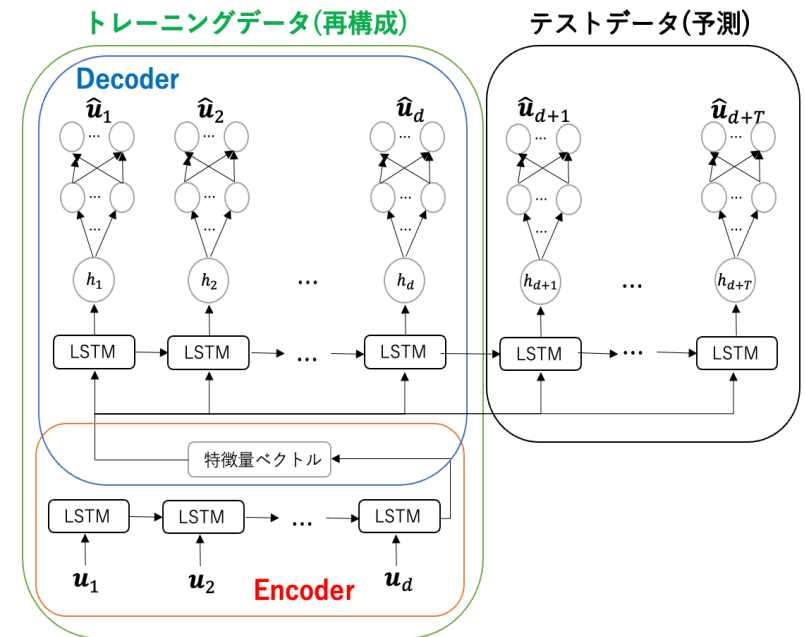


図1：提案モデル

パラメータの解釈可能性と異常値耐性

■ 使用データ

本研究ではHuman Mortality Databaseから日本の男女混合データを引用している。
年齢範囲：0歳～99歳、歴年範囲：1950年～2018年（トレーニング：1950年～1999年、テスト：2000年～2018年）

■ パラメータの解釈可能性

1950～2018年で再構成をして、SVDの時系列項 κ_t とLSTM-AEの出力 h_t は図2のようになる。 κ_t は減少関数、 h_t は増加関数で表現した『**医学的進歩による死亡率改善ファクター**』と解釈することができる。

■ 異常値耐性

SVDでは2010年付近で κ_t にスパイクが存在するが、これは東日本大震災による影響であると考えられる。しかし、 h_t では滑らかな曲線になっており、異常値耐性を確認できる。また、LSTM-AEの非線形表現能力のため h_t の単純化にもかかわらず、表1に示すようにSVDに再構成精度は劣ることなく学習可能であることが分かる。死亡率の長期予測の目的からは、一過性の影響を受けにくい異常値耐性を持つLSTM-AEの方が適していると言える。

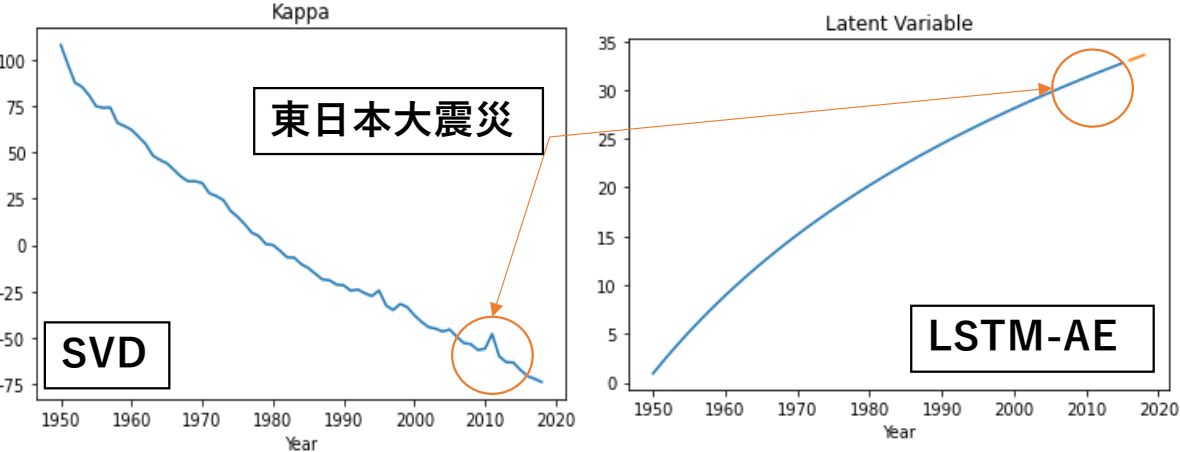


図2:(左)SVDの κ_t (右)LSTM-AEの h_t

表1:再構成精度(1950-2018)

再構成精度	LSTM-AE	SVD
MSE	0.042510475	0.076322891
MAE	0.003033394	0.009709561

外挿結果

■ 予測結果-対数死亡率の形状・予測精度

トレーニングデータで再構成を行い、テスト期間で予測精度を見る。提案モデルのようなNNモデルでは重み行列の初期値(シード)に対して出力結果が変わってしまうことから、10個のシード初期値で出力を行い、その平均を予測精度とする。

表2より、LSTM-AEの予測精度の方が、従来のLCモデルの予測精度を大きく上まっていることが分かる。特に図3より、対数死亡率の形状が崩れておらず、解釈及び説明が可能な出力結果を得ることができた。

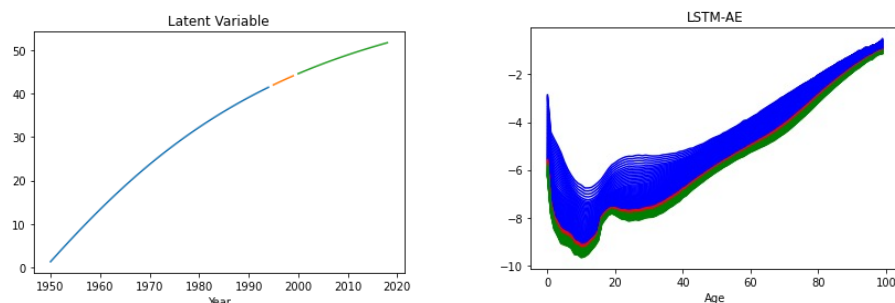


図3:(左) h_t の外挿,(右)対数死亡率の形状

表2:予測精度(2000-2018)

予測精度	LSTM-AE(平均)	LC
MSE	0.138292361	0.195655685
MAE	0.029847343	0.065635224

■ 長期外挿(50年先予測)

1950~2018年のデータを利用し、50年先の予測を行う。図4より、LCの結果では κ_t が直線的に減少しており、死亡率が非現実的に低下しすぎる可能性がある。

LSTM-AEの将来予測では、死亡率改善度はLCに比して緩やかに変化するため、**より自然な予測結果が得られる。**

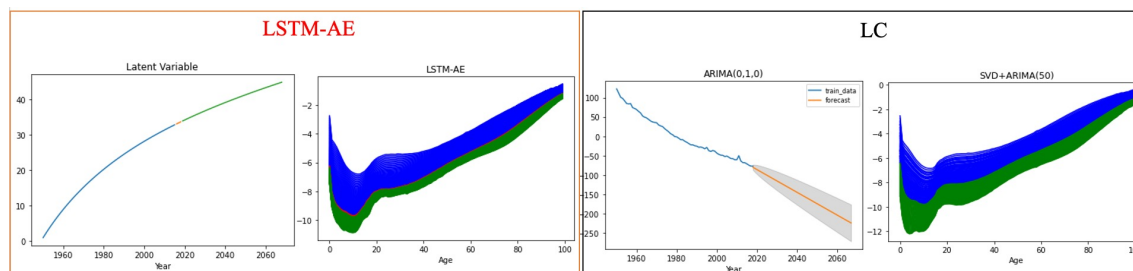


図4:長期外挿(左)LSTM-AE,(右)LC

まとめと課題

■ 初期値頑健性

NN特有の弱点としてシード初期値に対して出力結果が変わってしまうことがある。本研究ではシード値10個の出力結果の平均を予測精度として扱っているが、その出力結果のブレを箱髭図で表現すると図5のとおりであり、提案モデルの初期値頑健性を確認することができる。

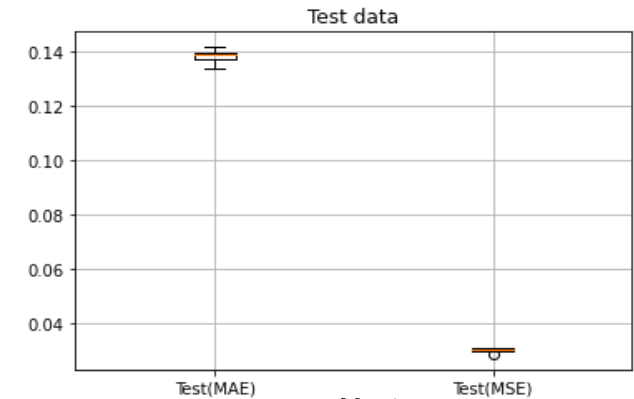


図5:箱髭図

■ まとめ

- ・ 本研究では初期値頑健性と解釈可能性のある1段階推定のNNによる死亡率予測モデルを提案した
- ・ LCよりも精度が改善されてかつ対数死亡率の形状が崩れることのない推定が可能であることを数値実験で確認できた
- ・ 地震等の一時的外乱要因の影響を受けにくい形で死亡率を推定することができ、今後のpost-COVID-19におけるリスク管理に適した異常値耐性を持つことが確認できた

■ 課題

- ・ 日本以外のデータや男女別データで精度を確認
- ・ 複数母集団化・コホート化に向けたモデルの拡張