Vol.2022-CVIM-228 No.18
2022/1/28

クラス情報を考慮したEnergy-based Modelによる時系列予測手法の提案

山縣 将貴[1,a)]　内田 誠一[1,b)]　早志 英朗[1,c)]

概要：本研究では，クラス情報を考慮した Energy-based Model (EBM) による時系列予測手法を提案する．対象の時系列情報がクラスの違いによる分岐構造を持つ場合，一意的な予測は不可能である．このような場合，分岐構造を学習するクラス情報を考慮した時系列予測が有用である．提案手法では，Softmax を用いた一般的な識別モデルが EBM として解釈可能であることを利用し，クラス予測と EBM による時系列予測を単一のモデルで同時に行う．これにより，入力系列がどのクラスに分岐する可能性を持っているのかを考慮した時系列予測を可能とする．

1. はじめに

時系列予測は重要であるものの，複雑な軌跡や非マルコフ性といった不確実性を高める要因を含むため非常に困難なタスクである．そのため，不確実性をモデル化することで予測精度を向上させる多くの試みがなされている．しかし，対象の時系列情報が分岐構造を持つ場合，一意的な予測は不可能である．筆記予測において数字の ‘0’ と ‘6’ を例に挙げると，筆記序盤はほぼ同様の軌跡である一方，筆記中盤以降は各数字に分岐し，大きく軌跡が変化する．このような場合，筆記序盤に ‘0’ と ‘6’ のどちらの数字が書かれるのかを確定することは不可能である．

このように，長期に渡る一意予測は不可能であるため，分岐構造を考慮した予測を行う必要がある．分岐構造を自動的に学習し複数の可能性を予測できるモデルは複数の視点から有用である．第一に，分岐点までは正確な予測結果が得られる．第二に，必要に応じて複数の予測結果を獲得できる．第三に，ある時刻以降に分岐点がないことが判明すれば信頼性の高い予測結果となる．

分岐構造を考慮した時系列予測は，現時刻までの時系列が与えられた条件下でのクラス予測とクラスごとの将来系列予測を同時に行う問題に帰着される．入力系列に対して複数クラスの予測確率が高い場合，予測系列がそれぞれのクラスに基づいた変化をしていくことで分岐が生じる．したがって，クラス予測の結果から入力系列が将来分岐する可能性を判断したうえで時系列予測を行うことで，分岐構造を考慮した時系列予測が可能となる．例えば，Yamagata ら [1] は入力された時系列のクラス予測と同時に，クラスごとの Gaussian mixture model（GMM）によって将来の予測分布を推定し，各 GMM の予測分布を統合することで分岐構造を考慮した時系列予測を実現している．Deo と Trivedi [2] は，予測対象とその周辺の自動車の軌跡を畳み込むことで得た特徴表現を用いて 6 クラスに分類した自動車の行動パターンを推定するとともに，自動車の軌跡をクラスごとに予測するモデルを提案している．

以上の背景から分かるように，分岐構造を考慮した時系列予測を行うにあたってクラス予測の confidence の精度が重要となる．ある予測クラスに対する confidence が低い場合，本来であれば分岐する可能性があったとしてもそれを考慮できないため，分岐構造を推定することができない．例えば，‘0’ と ‘6’ のどちらにも分岐する可能性があるにも関わらず，どちらか一方だけを予測してしまった場合，すなわち，どちらか一方の confidence が圧倒的に高くなってしまった場合には本来存在する分岐構造を見逃してしまうことになる．反対に，‘0’ であることは明らかであるにも関わらず，‘0’ と ‘6’ のどちらも同程度に予測してしまった場合，すなわち，どちらの confidence も 0.5 と予測してしまった場合には本来存在しない分岐構造を誤って推定してしまうことになる．このように，分岐構造を考慮するためにはクラス予測の confidence の精度が重要である．

一般的な深層学習系の手法では，この confidence の精度が問題になることが指摘されている．特に，クラス予測の学習に一般的に用いられているクロスエントロピーでは予測される confidence の精度が低くなってしまうという問

[1] 九州大学
Kyushu University
[a)] masaki.yamagata@human.ait.kyushu-u.ac.jp
[b)] uchida@ait.kyushu-u.ac.jp
[c)] hayashi@ait.kyushu-u.ac.jp

ⓒ 2022 Information Processing Society of Japan

題がある．これは，深層学習系の手法がほぼすべての学習データを正解できるようになった後もクロスエントロピーを下げることができ，過剰に高い confidence を出力するように適合してしまうためと考えられている [3].

このような問題を解決するためには confidence calibration を行うことが重要である．Confidence calibration とは，モデルの出力する confidence がその出力結果の正確さの尤度を表すように較正することである．confidence calibration を行う手法は多数提案されている [3], [4], [5]. 例えば，confidence を出力する際に通常の softmax ではなく，温度付き softmax を用いることで confidence の精度を改善できることが報告されている [3]．また，Grathwohl ら [5] は softmax を用いた識別モデルが生成モデルの一種である energy-based model（EBM）として解釈可能であることを示し，識別モデルを識別と生成のハイブリッドモデルとして扱う joint energy-based model（JEM）を提案することで confidence の精度を改善している．

分岐構造を考慮した時系列予測において，我々は識別・生成のハイブリッド学習により confidence calibration を行う JEM の方法論に着目する．JEM においてデータ生成を行っていた EBM を将来の系列予測に応用することで，分岐構造を考慮した時系列予測と confidence の改善を同時に達成できることが期待される．

本研究では，クラス情報を考慮した EBM による時系列予測手法を提案し，時系列予測の一例として手書き数字の筆記予測タスクに適用する．提案手法では，クラス予測に用いられる softmax が EBM と解釈できることを時系列を入力とする識別モデルに応用する．これにより，任意の識別モデルをもとにしてクラス事後確率と筆跡の予測分布の同時分布を単一モデルとして学習できる．これを活用することで，EBM による時系列予測と confidence の改善を同時に達成することを目指す．

2. 関連研究

2.1 分岐構造を考慮した時系列予測

分岐構造を考慮した時系列予測手法はいくつか提案されている．Lee と Kitani [6] は，マルコフ決定過程を用いて分岐構造を考慮した人間の行動を予測している．また，クラス情報と分岐構造を同時に考慮して時系列を予測する試みも行われている．Yamagata ら [1] は，は入力系列のクラス予測と同時にクラスごとの GMM によって将来の予測分布を推定し，各 GMM の予測分布を統合することで分岐構造を考慮した時系列予測を実現している．Pool ら [7] は，自動車の行動パターンごとに線形力学系を組み合わせることによって分岐を表現した自動車の行動予測モデルを提案している．Tang と Salakhutdinov [8] は，行動パターンを明示的にクラス分けすることなくマルチモーダルな行動予測を行うモデルを提案している．

提案手法のベースとなる EBM は予測分布に関してその形状やパラメータの仮定を必要としない．また，EBM は任意のニューラルネットワークに適用可能であるという利点を持つ．提案手法は，予測分布に関する仮定の不必要性と softmax を含む任意の時系列識別モデルへの適用可能性を兼ね備えている点でこれらの手法と異なる．

2.2 Energy-based model

Energy-based model（EBM）はエネルギー関数と呼ばれる正規化されていない負の対数確率をモデル化した生成モデルである．EBM における確率密度関数 $p_{\boldsymbol{\theta}}(\boldsymbol{x})$ はエネルギー関数 $E_{\boldsymbol{\theta}}(\boldsymbol{x})$ を用いて以下のように定式化される．

$$p_{\boldsymbol{\theta}}(\boldsymbol{x}) = \frac{\exp\left(-E_{\boldsymbol{\theta}}(\boldsymbol{x})\right)}{Z(\boldsymbol{\theta})} \tag{1}$$

ここで，$Z(\boldsymbol{\theta})$ は分配関数と呼ばれる $\boldsymbol{x}$ に関する正規化定数であり，$Z(\boldsymbol{\theta}) = \int \exp\left(-E_{\boldsymbol{\theta}}(\boldsymbol{x})\right)\mathrm{d}\boldsymbol{x}$ である．一般的に，$Z(\boldsymbol{\theta})$ は $\boldsymbol{x}$ に関する積分を含んでいるため計算できない．そのため，$Z(\boldsymbol{\theta})$ の推定にはマルコフ連鎖モンテカルロ法（MCMC）や stochastic gradient Langevin dynamics（SGLD） [9] が用いられる．

EBM は画像生成等の生成タスクにおいて広く用いられている．Arbel ら [10] は EBM を基にした生成モデルを提案しており，同程度のモデルサイズの GAN と比べて高品質なサンプルを生成できることを示した．Grathwohl ら [5] はクラス予測に一般的に用いられている softmax が EBM として解釈可能であることを示し，画像識別と生成のハイブリッドモデルである JEM を提案した．提案手法は JEM を時系列予測に応用したモデルと位置付けられる．

3. クラス情報を考慮した EBM による時系列予測手法

3.1 提案手法の定式化

提案手法は時刻 t までの D 次元系列 $\boldsymbol{x}^{(1:t)} = (\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(\tau)}, \ldots, \boldsymbol{x}^{(t)}), \boldsymbol{x}^{(\tau)} \in \mathbb{R}^D$ の入力に対して，$\boldsymbol{x}^{(1:t+1)}$ の属するクラス $y \in \{1, \ldots, K\}$ の事後確率 $p_{\boldsymbol{\theta}}(y \mid \boldsymbol{x}^{(1:t+1)})$ と時刻 $t+1$ の座標分布 $p_{\boldsymbol{\theta}}(\boldsymbol{x}^{(t+1)} \mid \boldsymbol{x}^{(1:t)})$ を予測する．ここで，K はクラス数，$\boldsymbol{\theta}$ は分布のパラメータである．

まず，クラス事後確率をロジット関数 $f_{\boldsymbol{\theta}} : \mathbb{R}^{D\times(t+1)} \to \mathbb{R}^K$ と softmax 関数を用いて以下のようにモデル化する．

$$p_{\boldsymbol{\theta}}(y \mid \boldsymbol{x}^{(1:t+1)}) = \frac{\exp\left(f_{\boldsymbol{\theta}}(\boldsymbol{x}^{(1:t+1)})[y]\right)}{\sum_{y'} \exp\left(f_{\boldsymbol{\theta}}(\boldsymbol{x}^{(1:t+1)})[y']\right)} \tag{2}$$

ここで，$f_{\boldsymbol{\theta}}(\boldsymbol{x}^{(1:t+1)})[y]$ は $f_{\boldsymbol{\theta}}(\boldsymbol{x}^{(1:t+1)})$ の第 y 次元，つまり，クラス y に対応するロジットである．

次に，$\boldsymbol{x}^{(t+1)}$ と y の条件付き同時分布 $p_{\boldsymbol{\theta}}(\boldsymbol{x}^{(t+1)}, y \mid \boldsymbol{x}^{(1:t)})$ を EBM を用いてモデル化する．このとき，式 (1) と式 (2) の類似性に着目し，$E_{\boldsymbol{\theta}}(\boldsymbol{x}^{(1:t+1)}, y) = -f_{\boldsymbol{\theta}}(\boldsymbol{x}^{(1:t+1)})[y]$ とお

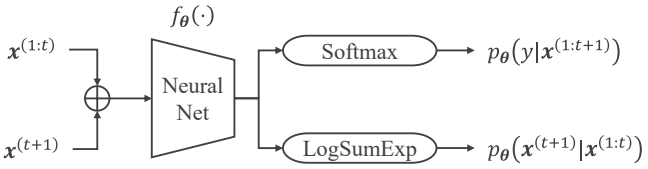


図 1: 提案手法の構造．$\bigoplus$ は系列の結合を表す．

いてロジット関数をエネルギー関数として再利用すれば，

$$p_\theta(\boldsymbol{x}^{(t+1)}, y \mid \boldsymbol{x}^{(1:t)}) = \frac{\exp\left(-E_\theta(\boldsymbol{x}^{(1:t+1)}, y)\right)}{Z(\boldsymbol{\theta})} = \frac{\exp\left(f_\theta(\boldsymbol{x}^{(1:t+1)})[y]\right)}{Z(\boldsymbol{\theta})} \tag{3}$$

と表される．また，式 (3) を y について周辺化すると

$$p_\theta(\boldsymbol{x}^{(t+1)} \mid \boldsymbol{x}^{(1:t)}) = \sum_y p_\theta(\boldsymbol{x}^{(t+1)}, y \mid \boldsymbol{x}^{(1:t)}) = \frac{\sum_y \exp\left(f_\theta(\boldsymbol{x}^{(1:t+1)})[y]\right)}{Z(\boldsymbol{\theta})} \tag{4}$$

が得られる．なお，右辺はエネルギー関数として

$$E_\theta(\boldsymbol{x}^{(t+1)}) = -\log \sum_y \exp\left(f_\theta(\boldsymbol{x}^{(1:t+1)})[y]\right) \tag{5}$$

を式 (1) に代入した EBM に相当する．提案手法では，式 (2) と (4) をもとにクラス確率予測と座標予測を行う．

3.2 提案手法の構造と最適化

提案手法の構造を図 1 に示す．任意のニューラルネットワークで構成されたロジット関数 $f_\theta(\cdot)$ は入力として $\boldsymbol{x}^{(1:t+1)}$ を受け付ける．$f_\theta(\cdot)$ の出力に softmax を適用することで $p_\theta(y \mid \boldsymbol{x}^{(1:t+1)})$ を，LogSumExp を適用することで $p_\theta(\boldsymbol{x}^{(t+1)} \mid \boldsymbol{x}^{(1:t)})$ を得る．ここで，LogSumExp は式 (5) の右辺の演算 $\log \sum_y \exp(\cdot)$ を表す．なお，$\boldsymbol{x}^{(t+1)}$ の初期値は一様分布からのサンプリングとし，その後は式 (5) で得られるエネルギーに SGLD を適用して逐次更新をする．

提案手法では式 (3) の同時分布の対数尤度である

$$\log p_\theta(\boldsymbol{x}^{(t+1)}, y \mid \boldsymbol{x}^{(1:t)}) = \log p_\theta(\boldsymbol{x}^{(t+1)} \mid \boldsymbol{x}^{(1:t)}) + \log p_\theta(y \mid \boldsymbol{x}^{(1:t+1)}) \tag{6}$$

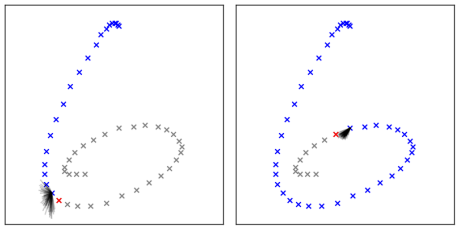
を目的関数として最適化を行う．右辺の第一項は $\boldsymbol{x}^{(t+1)}$ の座標分布の最適化，第二項はクラス確率分布の最適化を行う．また，右辺の第一項は $Z(\boldsymbol{\theta})$ を含むため，前述の通り SGLD を用いて最適化する．一方，右辺の第二項は式 (2) に示す通り，softmax と同一であるため通常の識別モデルと同じくクロスエントロピーを用いて最適化する．

4. 実験

4.1 実験条件

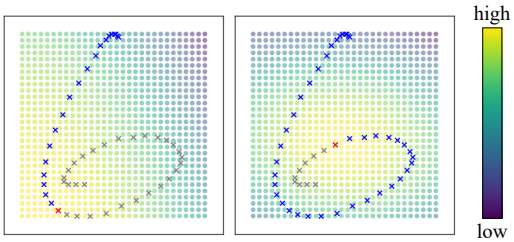
本実験の目的は，提案手法による時系列予測可能性の検証と座標分布学習の有無による confidence calibration への影響を検証することである．ただし，本実験は予備的結果にとどまっており，$t = 20$ と $t = 40$ まで入力した場合の 2 通りについて検証した．実験では，図 1 に示す提案手法の $f_\theta(\cdot)$ としてスキップコネクション付きの 3 層 LSTM を用いた．なお，各 LSTM の隠れ層は 400 次元とした．また，データには手書き数字 ‘0’ から ‘9’ の 10 クラス分の筆跡を二次元座標系列で表したデータセットである UNIPEN (Train-R01/V07)[11] を用いた．なお，本データセットにはクラスインバランスが存在し，各クラスに約 1,300 サンプルが用意されている．実験では訓練用に 70%，検証用に 10%，テスト用に 20%にランダムに分割して用いた．

実験では，confidence calibration の評価指標として Expected Calibration Error（ECE）[12] を使用した．ECE はモデルの出力する confidence とその出力の正答率の誤差を表す指標であり，ECE が小さいほど信頼性の高い確率予測であることを示す．また, confidence calibration のベースラインとして提案手法と同じ $f_\theta(\cdot)$ を用いた $p_\theta(\boldsymbol{x}^{(t+1)} \mid \boldsymbol{x}^{(1:t)})$ の推定を行わない一般的な識別モデルを用いた．



(a) $t = 20$ まで入力　(b) $t = 40$ まで入力

図 2: 提案手法による予測結果の例．黒色の実線は予測結果，青色の × 印は入力系列，赤色の × 印は正解座標，灰色の × 印は将来の系列を表す．



(a) $t = 20$ まで入力　(b) $t = 40$ まで入力

図 3: 提案手法によって推定された図 2 に示す予測結果のエネルギー分布．

4.2 予測結果の定性評価

図 2 に提案手法による予測結果の例を示す．図より，$t = 20$ まで入力した場合は正解座標から外れた点を予測してしまっている一方で，$t = 40$ まで入力した場合には概ね正解に近い座標を予測できたことが確認できる．予測結果全体を通じて，$t = 20$ まで入力した場合は比較的正確に予

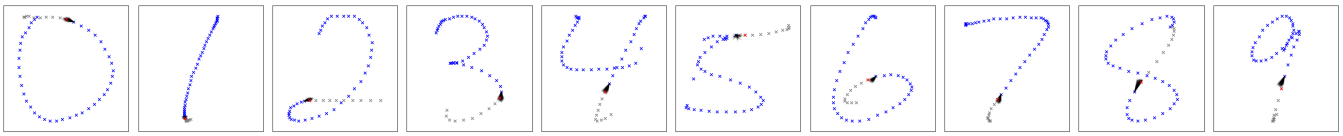


図 4: $t = 40$ まで入力した際の提案手法による各クラスのサンプルに対する予測例．実線と × 印の意味は図 2 に準ずる．

表 1: 提案手法とベースラインに $t = 20$ と $t = 40$ まで入力した場合の ECE.

Model	$t = 20$	$t = 40$
Ours	0.902	**0.812**
Baseline	**0.764**	0.868

測できた例もある一方で，全体的に予測が不安定である傾向が見られた．これは，$t = 20$ の時点では様々なクラスへ分岐する可能性によって予測の不確実性が高まり，適切な予測分布を推定できなかったためであると考えられる．

図 2 で示した予測結果のエネルギー分布を図 3 に示す．図より，$t = 20$ 及び $t = 40$ まで入力したどちらの場合でも正解座標の近辺を中心とするエネルギー分布が推定されていることを確認できる．エネルギー分布の分散は $t = 20$ の時よりも $t = 40$ の時のほうが小さくなっており，より精密な推定が行われていると言える．

また，$t = 40$ まで入力した際の各クラスの予測例を図 4 に示す．図より，$t = 40$ まで入力した場合は全てのクラスにおいてほぼ正確な座標予測が行えている．以上の結果から，提案手法による時系列予測の可能性が示唆された．

4.3 Confidence calibration の評価

表 1 に提案手法とベースラインの ECE を示す．表より，提案手法では $t = 20$ まで入力した場合と比較して $t = 40$ まで入力した場合のほうが ECE が 0.090 ポイント小さくなっており，confidence の精度が向上していることを確認できる．一方で，ベースラインでは $t = 20$ まで入力した場合と比較して $t = 40$ まで入力した場合のほうが ECE が 0.104 ポイント悪化していることを確認できる．また，入力長ごとに提案手法とベースラインの ECE を比較すると，$t = 20$ まで入力した場合は提案手法の ECE がベースラインの ECE を上回った一方で，$t = 40$ まで入力した場合は提案手法の ECE がベースラインの ECE を下回った．

このような結果を示したのは，座標値の発生頻度を表す座標分布の推定の有無が影響していると考えられる．提案手法では座標分布の最適化とクラス識別を同時に行っている．したがって，提案手法では予測した座標値の発生頻度を考慮したうえで識別を行うために，confidence の精度が改善したと考えられる．

一方で，$t = 20$ の場合に提案手法の ECE がベースラインを下回ったのは，座標分布の推定が十分に行えていなかったためであると考えられる．座標分布が正しく推定できていない場合，入力系列の発生頻度を誤って推定してしまうことになる．図 3 に示すように，$t = 20$ まで入力した場合はエネルギー分布の分散が大きくなっており，これが confidence calibration に悪影響を及ぼしたと考えられる．

5. まとめ

本研究では，分岐構造を考慮した予測をするうえで重要な confidence の精度を改善しつつ時系列予測を行うことを目的に，EBM に基づく時系列予測モデルを提案した．実験の結果，EBM に基づく提案手法によって時系列予測が可能であることを確認し，また，入力長が長い場合は座標分布を学習することによって confidence の精度が改善されることを確認した．今後の課題としては，予測精度の定量評価および汎用的な予測を可能とするために 1 時刻先だけではなく時系列全体を EBM によって一度に予測するよう拡張することが挙げられる．

参考文献

[1] Yamagata, M., Hayashi, H. and Uchida, S.: Handwriting Prediction Considering Inter-Class Bifurcation Structures, *ICFHR* (2020).
[2] Deo, N. and Trivedi, M. M.: Convolutional Social Pooling for Vehicle Trajectory Prediction, *CVPR Workshops* (2018).
[3] Guo, C., Pleiss, G., Sun, Y. and Weinberger, K. Q.: On Calibration of Modern Neural Networks, *ICML* (2017).
[4] Seo, S., Seo, P. H. and Han, B.: Learning for Single-Shot Confidence Calibration in Deep Neural Networks Through Stochastic Inferences, *CVPR* (2019).
[5] Grathwohl, W., Wang, K.-C., Jacobsen, J.-H., Duvenaud, D., Norouzi, M. and Swersky, K.: Your Classifier is Secretly an Energy Based Model and You Should Treat it Like One, *ICLR* (2020).
[6] Lee, N. and Kitani, K. M.: Predicting Wide Receiver Trajectories in American Football, *WACV* (2016).
[7] Pool, E. A. I., Kooij, J. F. and Gavrila, D. M.: Using Road Topology to Improve Cyclist Path Prediction, *Intelligent Vehicles Symposium* (2017).
[8] Tang, C. and Salakhutdinov, R. R.: Multiple Futures Prediction, *NeurIPS* (2019).
[9] Welling, M. and Teh, Y. W.: Bayesian Learning via Stochastic Gradient Langevin Dynamics, *ICML* (2011).
[10] Arbel, M., Zhou, L. and Gretton, A.: Generalized Energy Based Models, *ICLR* (2021).
[11] Guyon, I., Schomaker, L., Plamondon, R., Liberman, M. and Janet, S.: UNIPEN Project of On-line Data Exchange and Recognizer Benchmarks, *ICPR* (1994).
[12] Naeini, M. P., Cooper, G. F. and Hauskrecht, M.: Obtaining Well Calibrated Probabilities Using Bayesian Binning, *AAAI* (2015).

ⓒ 2022 Information Processing Society of Japan