

ニューラルネットによる系列モデルの最近の研究の紹介

三原 千尋 総合研究大学院大学 統計科学専攻 博士課程(5年一貫制)5年

系列モデル

系列データ $x_t (t = 1, 2, \dots, T)$ を順次に入力すると、その都度その時点までの入力を反映した特徴 h_t を出力するようなモデル

上記のようなモデルを系列モデルとよぶことにする。特にニューラルネットによる系列モデルは様々な系列データに対するタスクをニューラルネットで解く際に重要なコンポーネントとなるが、如何にその時点までの入力を反映した特徴を生成するかによって、基本的なネットワーク構造は表1の3種類に大別されると考えられる。

再帰的ニューラルネット (RNN)	前時点の特徴 h_{t-1} と現時点の入力 x_t から現時点の特徴 h_t を生成する。過去の情報が特徴 h_{t-1} に委ねられるので情報の流れの制御が重要になり、ゲート機構をもつ LSTM や GRU が広く用いられている。
畳込みニューラルネット (CNN)	現時点までの一定範囲の入力 $x_{(t-T):t}$ を直接参照して現時点の特徴 h_t を生成する。効率的に広範囲のステップを参照する汎用的なネットワーク構造として TCN (Temporal Convolutional Networks) [Bai2018] が提案されている (図1)。
その他	特徴の再帰も、入力の畳込みもせずにその時点の特徴を生成する。セルフアテンション (Vaswani et al., 2017) は位置エンコードを用いてこれを実現している。

表 1: ニューラルネットによる系列モデルの基本的な構造の分類

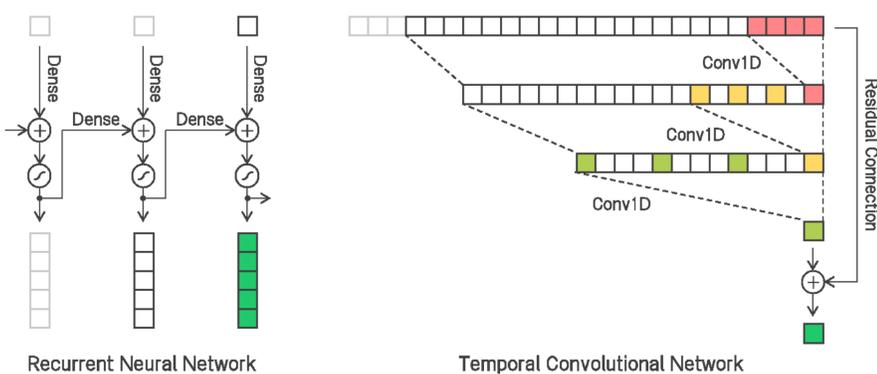


図 1: RNN と TCN の模式図 (Dense は全結合、緑で塗りつぶした四角は現時点の特徴)

過去の入力を直接参照する TCN は長い文脈をもつタスクで優れた性能を示す傾向があり、参照範囲を固定しない RNN はより転移学習に柔軟性をもつと考えられる [Bai2018]。他方、QRNN (Bradbury et al., 2017) など、RNN と CNN 双方の特徴抽出の性質を併せもつモデルも多数提案されている。以下では特に多岐にわたるタスクで既存モデルの性能を上回った TrellisNet [Bai2019] について紹介する。

TrellisNet[Bai2019]

TrellisNet は時刻 $t+1$ の $i+1$ 層目の出力 $z_{t+1}^{(i+1)}$ が以下で表される。ここで、 W_1, W_2 は何層目かに依らない重み行列であり、 f は適当な非線形変換による活性化である。また、 $z_t^{(0)} \equiv 0$ とする。

TrellisNet

$$\tilde{z}_{t+1}^{(i+1)} = W_1 \begin{pmatrix} x_t \\ z_t^{(i)} \end{pmatrix} + W_2 \begin{pmatrix} x_{t+1} \\ z_{t+1}^{(i)} \end{pmatrix}$$

$$z_{t+1}^{(i+1)} = f(\tilde{z}_{t+1}^{(i+1)}, z_t^{(i)})$$

TrellisNet は畳込み層が繰り返される点は多層 TCN と同様であるが、各層の重みが共有されている点、各層に入力 x_t が injection される点が多層 TCN と異なる (図2)。特にこの injection があることにより、TrellisNet は「現時点までの M ステップを参照する L 層 RNN が常に $M+L-1$ 層の TrellisNet で表現できる」という性質をもつ (図3)。

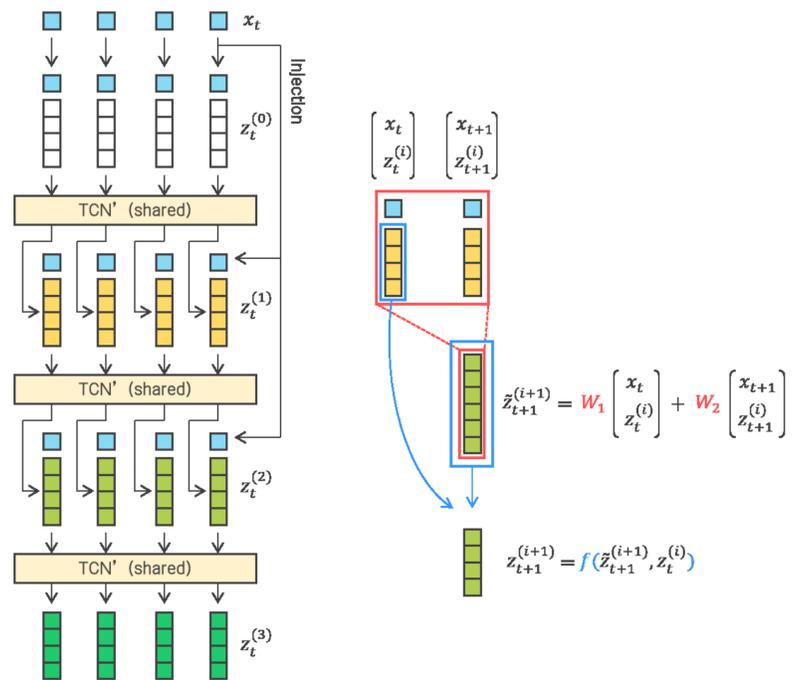


図 2: TrellisNet の模式図。各層の操作は現時点までの2ステップを1回だけ畳込む TCN といえる (が TrellisNet は活性化 f の適用時に $z_t^{(i)}$ も利用する自由度がもたされている)。実際には f には LSTM cell に基づく gated activation が採用されている。

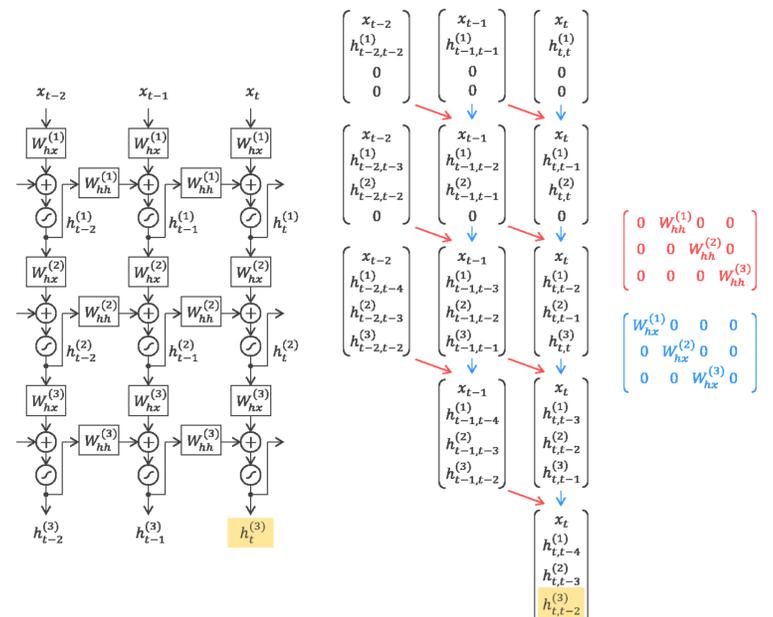


図 3: 現時点までの3ステップを参照する3層RNN(左)とそのTrellisNet表現(右)。但し $h_{t,t}^{(i)}$ は時点 t に開始した RNN の時点 i における t 層目の出力であり、赤と青の矢印は同じ文字色の重み行列を適用することを意味する。

実践的には、TrellisNet は正則化や最適化に RNN と CNN の双方からアイデアを取り入れられることが利点になっており、実際、LSTM の DropConnect (Merity et al., 2018) や CNN の auxiliary losses (Lee et al., 2015; Xie & Tu, 2015) などを取り入れ、Penn Treebank、WikiText-103、その他長期記憶性能テストという多岐にわたるタスクでそれぞれの既存モデルの性能を刷新している。今後の展望には f の構造探索やセルフアテンションの取り込みなどが挙げられている。

参考文献

- [Bai2018] Shaojie Bai, J. Zico Kolter, Vladlen Koltun. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling, *arXiv:1803.01271*, 2018. <https://arxiv.org/abs/1803.01271>.
- [Bai2019] Shaojie Bai, J. Zico Kolter, Vladlen Koltun. Trellis Networks for Sequence Modeling, In *International Conference on Learning Representations (ICLR)*, 2019. <https://arxiv.org/abs/1810.06682>.