

V-SOINN と相互作用する SOINN

日大生産工 ○小林 拓也 日大生産工 山内 ゆかり

1. まえがき

データの可視化は、観察者にデータを見やすくすることでデータ分析において重要な役割を果たす。特に多次元データを可視化するには、データを人間にとって見やすい低次元空間へ写像することが必要とされている。また、膨大なデータにはノイズが含まれており、対処が必要である。山崎らが提案したSOINNは機械学習の手法の一つであり、高いノイズ耐性を持つ。しかし、SOINNは高次元データに対する性能が低いことで知られている。そこでHuiらはデータ点内のトポロジー関係を保持しながら低次元空間での投影を生成できるニューラルネットワークであるV-SOINNを提案し、人工データセットと実データセットで行った実験ではV-SOINNはPCA,MDS,t-SNE,NeuralGas,SOMを上回る可視ノードのクラス分け性能を示し、かつトポロジーを保持することができ、サンプル数が少ない場合にも正しい分類結果を生成することができたという報告がされている。

本研究では、V-SOINN空間の精度に関して、SOINNの高次元データに対する性能の低さの影響を下げるために、V-SOINN空間にノードの価値を導入し、その価値によってノードの追加・削除を行うことでV-SOINN空間の精度の向上を目指すことを提案し、MNISTとoptdigitsを用いた実験により提案手法との精度を比較し、得られた結果について報告する。

2. 従来研究

2.1 Adjusted SOINN

自己増殖型ニューラルネットワーク(SOINN)には複数の応用事例が存在するが、それらで共通して利用される最も基本的なアルゴリズムであるAdjusted SOINNについて説明する。Adjusted SOINNは一層構造で、パラメータ数は2つである。また、オンライン学習を用いた自己増殖的なネットワークの構成と、構成されたネットワーク構造に基づく入力分布の近似を可能にする。以下の図1にフローチャートを示す。

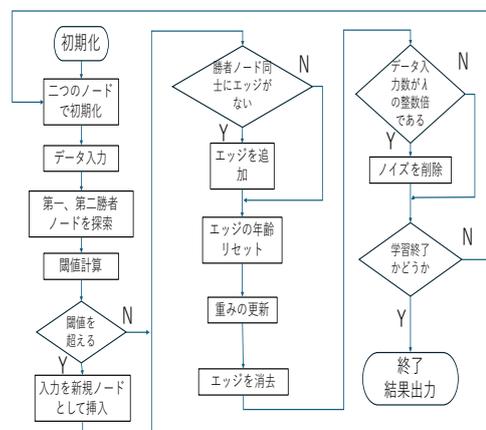


Figure 1

Adjusted SOINNのフローチャート

Adjusted SOINNではまず学習データからランダムに選択されたノードのみにする初期化が行われる。次に入力として学習データが与えられると各ノードとの距離を計算し、最も距離の近いノードである第一勝者ノードと、2番目に距離の近い第二勝者ノードが求められる。その際に、類似度閾値も求められる。類似度閾値は勝者ノードがそれぞれ連結しているノードがある場合には勝者ノードから最も遠いノードとの距離を、勝者ノードがそれぞれ連結しているノードがない場合には勝者ノードから最も近いノードとの距離が代入される。以下に閾値を求める2つの式を示す。

$$T_i \leftarrow \max_{c \in N_i} \|W_i - W_c\| \quad (1)$$

$$T_i \leftarrow \min_{c \in A \setminus \{i\}} \|W_i - W_c\| \quad (2)$$

入力と第一勝者ノード、第二勝者ノードとの距離のどちらかが各閾値を超えた場合、入力を新規ノードとして追加する。どちらも各閾値を超えない場合には、まず、2つの勝者ノード間にエッジがない場合にはエッジが追加される。次に、エッジが存在する場合には、そのエッジの年齢が0にリセットされ、第一勝者ノードにつながる全エッジの年齢が1増やされる。その後、式(3)、(4)により重みの更新を行う。

第1勝者ノードは式(3)により入力ノードから勝者ノードを引いたものを、そのノードが勝者ノードに選ばれた回数で割り、その値をもとの重みに加える。

次に、その他のノードは(4)の式のように、入力ノードから各ノードの値を引いたものに対して、各ノードが勝者ノードに選ばれた回数×100で割り、その値をもとのノードの重みに加える。以上により各ノードの重みの更新を行う。

$$\Delta W_{s1} \leftarrow \frac{1}{t_i}(\xi - W_{s1}) \quad (3)$$

$\Delta W_i \leftarrow \frac{1}{100t_i}(\xi - W_{s1}), \forall i \in N_{s1} \quad (4)$
--

重みの更新後、エッジ年齢が年齢閾値を超えたエッジを削除し、さらに削除したエッジで連結されていたノードのうちほかのすべてのノードとのエッジの接続がなくなったノードを削除する。また、入力回数が削除の周期である λ の倍数である場合に、接続されているエッジの数が一本以下であるノードをすべて削除する。この周期を学習が終了するまで繰り返す。

2.2 V-SOINN

V-SOINNは、高次元空間のトポロジカルなネットワーク構造を2次元空間でも維持させるために以下のようなことを行う。

高次元ネットワークで描画された新しいノード $a \in A$ に対して、2次元ネットワークに対応する可視ノード $a' \in A'$ を持つようにする。

具体的には、高次元空間ノードの重み w_a と対応低次元空間ノードの重み w'_a を、それぞれの平均と分散を使って正規化する。式は以下の通りになる。 σ 、 u はそれぞれ平均と分散を表す。

$$w_a = \frac{w_a - u}{\sigma}, w'_a = \frac{w'_a - u'}{\sigma'} \quad (5)$$

次に、低次元空間の重みを以下の式により更新する。

$$S(a, b) = (d(w_a - w_b) - d(w'_a - w'_b))^2 \quad (6)$$

$S_1(a) = \sum_{b \in A} S(a, b) \quad (7)$

各ノードの高次元空間と2次元空間のそれぞれのユークリッド距離の差を求め、それを距離差 $S_1(a)$ が最小になるように偏微分をする。その後以下の式のように低次元空間の重みを更新する。

$$w'_a \leftarrow w'_a - \eta_1 \frac{\partial S_1(a)}{\partial w'_a} \quad (8)$$

これを学習回数分おこなうことで、低次元空間のデータは高次元空間の関係性を写像していき、低次元空間でも高次元空間のトポロジカルなネットワーク構造を維持することができるようになる。

3. 提案手法

低次元空間の学習の中で、価値の概念を導入し、各ノードの価値を元に削除の判断を行い、それをもとにV-SOINN空間の精度の向上を目指す。価値として今回各ノードの他の全てのノードとの距離を利用する。従来研究で連結ノード数が1以下のノードはノイズの影響を受けているという判断をされ、ノードが削除されていた。そのため、疎なデータ空間上のノードは自然と削除されてしまうことがある。今回提案した価値はそのノードがどれだけ疎かを表すものであり、その価値が大きいノードに対する削除を減少させることにより、疎なノードを残すことで、V-SOINN空間における精度の向上を見込むことが出来ると考える。V-SOINN空間ではこの価値を導入した削除方法によって、SOINNに影響されずにV-SOINN空間の学習の安定化を図る。

4. 実験および検討

実験では、学習データ及びテストデータセットとして、MNISTとoptdigitsを利用する。SOINN空間と共にV-SOINN空間でも学習を行い、ノードの削除についてはSOINNに依存せずV-SOINN空間における価値をもとに削除を行う。

学習後のV-SOINN空間が従来研究と比べてどれほどのクラス分類精度が得られたかを提案手法と比較する。

5. まとめ

本研究では、V-SOINN空間にて価値を導入することにより低次元空間での学習の安定化を図ることを行った。SOINNの高次元に対する性能の低さを、V-SOINNを利用することで性能の低さに関わらず学習することを実践した。

参考文献

- 1) 山崎和博, 巻渕有哉, 申富饒, 長谷川修, SOINNとその実践, 日本神経回路学会誌 Vol.17, No.4(2010), 187-196
- 2) Hui Dou, Baile Xu, Furao Shen and Jian Zhao, "V-SOINN: A topology preserving visualization method for multidimensional data",