

欠損値推定のための複数SOM協調学習法

国際環境工学部 機械システム工学科 教授 岡田 伸廣

1. 欠損値とその推定

大量のデータを集めて、そこから各種の情報を取り出そうとする際、欠損値というものが問題となります。例えばワインの銘柄ごとに味、香りや成分などの計測値を集めたデータを作成し、そのデータに対する何らかの統計処理をしたいとします。ここで各銘柄を1つの実例とし、各実例が変数として各計測値を持つと考えます。このとき、いくつかの実例においてどれかの変数の計測が行われていなければ、その変数が欠損値となって、統計処理を行うことができなくなります。あるいは、複数の調査結果を統合しようとした際、ある調査では1種の変数についての計測が行われていなければ、その調査に含まれるすべての実例において、その変数が欠損値となります。そこで、なんらかの手法を用いてこの欠損値を推定する必要が生じます。

2. SOMを用いた欠損値推定

自己組織化マップ(Self-Organizing Map: SOM)は Kohonenによって提案された教師なし学習を行うニューラルネットワークの1種で、学習によって入力データ間の非線形な関係性を推定するのに用いることができます。そこで、このSOMを用いて欠損値を推定する手法があります。その際、欠損のない実例を学習用データとしてSOMに与えて変数の間の関連性を学習させ、学習後のSOMに他の変数の値から欠損値を推定させます。しかしながら、1つの実例が持つ変数の数が増加すると、どの実例においてもいずれかの変数が欠損している可能性が高くなり、学習に使用できる欠損のない実例が少數となって、精度の良い学習・推定を行えなくなってしまいます。

3. 複数SOMの協調

本研究室では、複数のSOMを協調させることによって、上記の問題を解決する手法について研究を行っています。各SOMはそれぞれ、全ての変数のうち一部の変数を無視するようになっています。図1において、各実例は6個の変数を含みますが、実例Bと実例Cにおいてそれぞれ、5番目の変数と4番目の変数が欠損値となっており、このままでは実

例Aしか欠損のない実例としてSOMの学習に使用することができません。しかしながら図2のように、変数5と6を無視して変数1~4についてのみ変数間の関連性を学習するSOMを用意すれば、そのSOMは実例Aと実例Bを用いて学習を行い、実例Cにおいて欠損している変数4の値を十分な精度で推定することができます。また別に図3のように、変数3と4を無視して学習を行うSOMを用意すれば、実例AとCを用いて学習を行い、実例Bにおいて欠損している変数5の値を推定できます。このようにすると、それぞれ無視する変数が異なるSOMを複数用意すれば、それらの組み合わせで推定することで、どの欠損値に対しても十分な精度で値の推定を行うことができるようになります。

実例	変数					
	変数 1	変数 2	変数 3	変数 4	変数 5	変数 6
実例 A	10	赤		50	爽快	100
実例 B	25	白		60	渋み	微甘
実例 C	20	ロゼ		80		甘

← 学習に使用できる
← 学習に使用できない
← 学習に使用できない

図1 欠損値を含むデータの例

実例	変数					
	変数 1	変数 2	変数 3	変数 4	変数 5	変数 6
実例 A	10	赤		50	爽快	100
実例 B	25	白		60	渋み	微甘
実例 C	20	ロゼ		80		甘

← 学習に使用できる
← 学習に使用できる
← 学習に使用できない

↑ 実例 A・B を用いて学習を行い、
実例 C の変数 4 の欠損値を推定

変数 1 ~ 4 のみを使用して、変数 5・6 を無視する SOM

図2 変数5・6を無視するSOMの利用

実例	変数					
	変数 1	変数 2	変数 3	変数 4	変数 5	変数 6
実例 A	10	赤		50	爽快	100
実例 B	25	白		60	渋み	微甘
実例 C	20	ロゼ		80		甘

← 学習に使用できる
← 学習に使用できない
← 学習に使用できる

↑ 実例 A・C を用いて学習を行い、
実例 B の変数 5 の欠損値を推定

変数 1・2・5・6 のみを使用して、変数 3・4 を無視する SOM

図3 変数3・4を無視するSOMの利用

4. 複数SOMの問題点と協調学習

複数SOMによる欠損値推定システムでは、1つの欠損値に対して複数のSOMが推定値を出力します。そのため、推定値を決定するためにはそれらを統合する必要があります。そこで、複数SOM推定値の分散が小さい場合のみ平均を用いて、分散が大きい場合にはもう一段階の学習を行ってから推定を行う等の手法や、単純に複数SOMの推定値の平均を用いる手法を考案して、シミュレーションによって各手法の比較検討を行いました。その結果、現時点では単純に推定値の平均を用いる手法が最も精度よく推定を行えるようだとの結論を得ました。表1に実際のワインに関するデータを欠損させた上で欠損値の推定を行った結果を示します。分散に基づいて二段階の学習を行う提案手法Iに比較して、単純な平均を用いる提案手法IIの方が小さな推定誤差となっています。また提案手法は単一のSOMを使用するなどの他の手法に比較して、いずれよりも推定誤差が小さくなっています。

また、SOMが精確な推定を行うためには、学習のための実例群を用いて多数回の学習を行う必要があります。一旦学習が済んでしまえば推定のための計算は極僅かとなりますが、その準備のための学習に相当な計算時間が必要なのです。そこで、学習を行う際にSOM同士で情報を共有することにより、学習時間を短縮する手法を開発しました。具体的には、他のSOMが学習を行った結果を次のSOMが学習を行うための初期値として使用することにより、必要な学習回数を減らします。図4に、協調学習によって学習回数を半分に減らした場合の欠損値推定誤差を示します。横軸は変数を欠損させた割合です。グラフより、学習回数を減らしても誤差の増大を抑えることができたことが分かります。

5. 今後の課題

今後は、学習段階での各SOM間での協調をさらに工夫して、より少ない学習回数でのより高精度の推定を実現したいと考えています。現在は、より協調の度合いが高いと考えられる各種の協調学習法についてシミュレーションを行ってその有効性の比較検討を行っています。また、現在用いている協調学習法では、本来対等であるべき各SOMの間に学習の順番による違いが生じているため、その違いを解消する方法についても検討を行っていきます。

表1 提案手法およびその他の手法を用いたワインデータ欠損値推定の相対誤差 [%]

	Data 1	Data 2	Data 3	Data 4	Average
Proposed method I	59.2	62.1	58.9	58.5	59.7
Proposed method II	53.1	56.9	54.5	53.0	54.4
Conventional method by a SOM	60.3	69.4	66.2	66.6	65.6
Method using the average	81.0	85.9	81.2	80.7	82.2
hot_deck	67.3	76.3	69.9	71.7	71.3

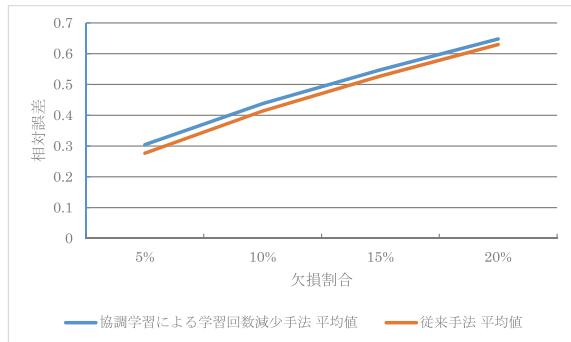


図4 協調学習を用いて学習回数を減らした場合の相対推定誤差

Profile



岡田 伸廣

Nobuhiro Okada

役職／教授

学位／博士(工学)

学位授与機関／九州大学

【連絡先】 n-okada@kitakyu-u.ac.jp

- 研究分野・専門 ロボット工学, システム工学
- 主要研究テーマ 3次元画像計測の応用とセンサシステム, 複数自己組織化マップの協調学習
- P R・その他 専門はロボット工学のうち、特に3次元画像計測を中心としたコンピュータビジョンに関する研究です。3次元画像計測のためのセンサシステムおよびそれに関連して、レーザ光スキャナの開発も行っています。3次元画像からの対象情報を用いて、人体をはじめとした柔軟対象の変形の計測手法についても研究を行っています。また、本稿で紹介した複数SOMの協調学習手法の開発およびその応用に関する研究や、壁面移動ロボットのための移動機構の開発など、ロボットおよびシステムに関して幅広く研究者を行っています。