

称号及び氏名	博士(学術) Suwardi
学位授与の日付	平成19年3月31日
論文名	「Multivariate Analysis and Information Visualization in Environmental Data Mining (多変量解析と可視化による環境情報のデータマイニング)」
論文審査委員	主査 小山 修平 副査 藤浦 建史 副査 村瀬 治比古

論文要旨

1. はじめに

インドネシアのような農業国では、農業計画や国土保全・整備に係わる様々な環境情報をできる限り、正確（一定の精度）で、かつ迅速に把握しなければならない。特に、本研究では、インドネシアにおいて農業に対して最も重大な環境要因である熱帯雨林地域の乾季において頻繁に発生している森林火災及び熱帯降雨パターン等の予測を中心に統計学（多変量解析）と可視化技術を活用として、より適切なデータマイニング技法を提起することを目的とする。

本論文は、序論（第1章）・第1パート（第2章～第3章）・第2パート（第4章～第6章）・結論（第7章）で構成されている。ここでは、インドネシアにおいて観測された様々な環境情報を用いて統計解析、可視化及びその考察に関する技法について、いくつかの新しい提案を行う。通常、観測（調査）に伴うデータ分析では、その環境における現象について、極論すれば、部分的な説明をしているにすぎない。その結果、観測者は多くの現象に関する複雑性を把握するために、異なった変数の観測データを大量に収集する必要がある。しかしながら、環境情報には、ノイズ（誤差）分布、冗長度を始め、非常に煩雑でしかも不可能に近い全変数観測が必要となり、適切なデータ分析技法は、ほとんど確立されていない。いくつかの空間的な現象の観測では、スケールが大きく異なったノイズパターンから生じる変動値を制御することは困難である。したがって、これらのデータの類型に関する探究は非常に重要であり、結果的に、あるデータから重要な知識や有用な情報を容易に得られる技法を提案することが大切である。一般に、次元の低減や可視化は、複雑なデータの表示または考察を可能にすることから複雑なデータを扱う際の標準的な技法の一つとなっている。本研究では、従来、個別的に用いられているこれらの技法を（すなわち、環境情報に関わるデータ分析のために多変量解析とデータマイニング技法を組み合わせ

て) 分析する方法を提案する。一般に多変量解析では、データ間の相互関係の把握や多変数の同時観測値を含んだデータを取り扱うことができる。一方、データマイニングの目的は、大量または複雑なデータの集まりから、通常の方法では見つけることができないような隠れた有意な情報を見出すことである。さらに、データを可視化して表示することは、データパターンの認識を容易にすることができる。本研究におけるデータ分析では、変数間の相互関係や依存、クラスタリング、分類、予測及び低次元化を行った。技法としては、主成分分析 (PCA)、判別分析 (DA) 並びに AMMI モデル、自己組織化マップ (SOM)、ニューラルネットワーク及びニューロファジィシステムを用いる。

ここでは、2つの環境情報に関するデータを分析するために、新たな解析技法の提案を行う。最初に取り扱う環境情報は、インドネシアのスマトラ島とカリマンタン島における多くの熱帯雨林地域で森林火災が生じるリスクに関して影響する空間的なホットスポット (熱点) についてのデータである。もう1つの環境情報はスラウェシ島で観測された日及び月ごとの雨水中に含まれる化学物質が記録された熱帯雨林地域の降雨データである。これらは、インドネシア内で実測されたデータであり、非常に多くのデータ数、多次元の変数及びデータ値が大きな範囲にあることから複雑なデータとなっている。したがって、このまま統計解析を行っても十分な知識を得ることは困難であり、ここでは、まず、データ情報 (変数) の低次元化をはかることになる。本研究における分析、結果並びに考察の大部分は第1パートと第2パートにおいて述べる。

2. 第1パートでは、

ホットスポットデータの取り扱いについて述べる。インドネシアの多くの熱帯雨林地域で森林火災が生じるリスクの一因となっているホットスポット分布の拡大から作成されたデータは、主として観測地域が広いことと、さまざまな分布特性を有し複雑である。この複雑さの解析とその考察は2つの部分で構成されている。特に、PCA と SOM がホットスポットデータに基づいて熱帯雨林地域で森林火災のリスクを分類して可視化するために用いられた。これら2つの技法 (PCA と SOM) によって、データ変数の低次元化やマッピング (写像) を行った。しかしながら、熱帯雨林地域で森林火災が生じるリスク評価について、より適切に表現 (可視化) するには、PCA より SOM の方が優れていることが分かってきた。SOM では、いくつかの種類に SOM ノードで森林火災が生じるリスク評価の分布をある程度分類することができた。しかしながら、SOM ノードの距離行列だけでは、分類した境界を明確に示すことができない場合もあり、分類境界が不明確なままでは、分析結果を解釈することが困難になる。したがって、SOM 分類について別の方法でより詳細な説明を必要とすることも分かってきた。熱帯雨林地域で森林火災が生じるリスクについて、より明確な分類技法について検討した。そのために、SOM ノードの彩色変化に対して判別分析 (DA) を適用し、SOM による分類をより明確にすることを試みた結果、DA を用いることによって、かなり明確に SOM 境界を示すことができた。したがって、SOM と判別分析 (DA) を組み合わせた技法は、データを可視化してクラスタリングできる非常に有用なものであり、熱帯雨林地域で森林火災が生じるリスクの分布を、よりの確に表すことを可能にした。

3. 第2パートでは、

熱帯雨林地域における降雨データの取り扱いについて述べる。第2パートは3つの部分で構成されている。第1部では、さまざまな時期や降雨継続時間の雨水中に含まれた化学物質の濃度を周期的に測定することは、雨水の汚染を評価する上で、たいへん重要なことを示した。ここでは AMMI モデルと SOM を降雨データの2つの分析要素（降雨の継続時間や時期：要素1と化学物質の内容：要素2）の相互作用を見出すために用いた。AMMI モデルは加法型と乗法型相互作用効果を分散分析に組み合わせており、その分析結果は、分析要素の相互作用効果がF検定（有意水準5%）で有意となり、かなりの相異を示すことができた。この分析技法では相互作用の要素1だけが重要となることから、要素2について相互作用係数を表示するために AMMI の散布図を使用できないことになる。しかし、SOM 分類を適用して雨水中に含まれる化学物質の濃度分布を評価すると、降雨の時期・継続時間に関わる雨水中の化学物質の相互作用パターンの相違を明らかにすることができた。第2部では、熱帯雨林気候の変動が熱帯雨林地域において不定期的に生じる降雨パターンの変化に影響を及ぼすことから日降雨変動を適切に予測するための解析技法を開発することは極めて難しい。ここでは、熱帯雨林地域における日降雨量を的確に推定するために、人工ニューラルネットワークとファジィ推論を組み合わせたニューロファジィ分類モデルを構築して、熱帯雨林地域における日降雨パターンの推定を試みた。その結果、学習後のより簡潔なルールを用いて日降雨量における変動パターンを高精度で推定できることを明らかにした。このように複雑なデータに比較的簡単なニューロファジィ分類モデルを適用するだけで、ある程度データマイニングが可能となることを示した。この技法による分析は、簡単なルールによって高精度で安定した結果を提供し、熱帯雨林地域の日降雨量における変動パターンについてより適切な考察を可能にした。

このパートの第3部では、前述のように、熱帯雨林地域における気候と降雨の非線形性は、熱帯雨林地域の降雨パターンの複雑さに反映し降雨量を推定することは極めて難しい。したがって、まず、雨季の降雨量の変動について、前述のファジィ推論モデルを用いて、よりの確な考察を可能にする簡単なファジィ分類ルールを作成した。このファジィルールに基づいて、降雨量の変動を予測するために適応型ニューロファジィ推論システムを構築し、平均2乗平方根誤差（RMSE）を用いて予測モデルの精度を評価した。その結果、本予測モデルを使用して熱帯雨林地域における雨季の降雨量の年変動を予測する場合、高精度で推定できることが分かった。結果的に、提案した予測モデルの妥当性も検証することができた。

4. おわりに

本研究では、インドネシアにおいて実測された環境情報データを対象に、多変量解析と情報可視化技法を組み合わせ、熱帯雨林地域におけるホットスポットデータのリスク評価、雨水中の化学物質の濃度分布、降雨パターンの変動及び降雨量の予測などに、より適切で有用なデータマイニング技法を提案することができた。これらの技法は、農業情報や環境情報などのデータのよりの確な取り扱いを可能にし、熱帯雨林地域の農業・環境計画並びに環境保全等にとって重要な情報を提供できることが分かった。

審査結果の要旨

自然環境や農業に関わる情報（いわゆる環境情報）などの収集は、人工衛星画像データに代表されるように、観測機器と計測技術の急激な進歩によって複雑かつ大規模化している。特に、適正な農業計画や環境保全・整備計画を立案する場合、収集された大規模な環境情報をできる限り一定の精度で迅速に分析・評価し、将来の変動を予測することが重要である。本研究で対象とする環境情報は、インドネシアのスマトラ島とカリマンタン島における熱帯雨林地域での森林火災のリスクに関わるホットスポット（熱点）データならびにスラウェシ島で観測された日および月降雨中に含まれる化学物質が記録された降雨データである。これらは、インドネシアで観測されたデータであり、大規模なデータ数、多次元の変数およびデータ値が大きな範囲にあることから非常に複雑なデータ構造となっている。一般に、このようなフィールドデータに対して通常の統計分析を適用しても膨大で複雑なノイズパターンを含んでおり、収集情報から必要なデータを抽出して活用することは極めて困難である。近年、通常の統計分析に対してファジィ推論・人工ニューラルネットワークなどのデータ処理ならびに自己組織化マップ（SOM）によるマッピング（写像）や可視化に代表されるデータマイニング技法の発展は著しく、さまざまな分野で活用されている。しかしながら、熱帯雨林地域の環境情報は、時系列的で、あまりにも広範囲な変動・変化の要因が多いことから、環境保全や農業計画に有効なデータとして利活用するためのデータマイニング技法は、ほとんど確立されていない。したがって、本研究では、データ解析として変数間の相互関係・依存、クラスタリング、分類、予測および低次元化を行うために主成分分析（PCA）、判別分析（DA）、AMMIモデル、自己組織化マップ（SOM）、ニューロファジィモデルおよびニューロファジィ推論システムを適切に組み合わせてインドネシアを対象に自然環境や農業上最も重大な環境要因である熱帯雨林地域において頻繁に発生する森林火災のリスク評価および降雨パターンの変動予測を明らかにするとともに環境情報にとってより適切なデータマイニング技法を提案することである。ここで得られた研究成果は次のとおりである。

1. インドネシアの多くの熱帯雨林地域で森林火災が生じるリスクに関係している熱点データは、主として観測領域が広く、さまざまな分布特性を有し複雑である。この複雑なデータを解析するために、従来の技法であるPCAと新たに導入したSOMによって、データ変数の低次元化やマッピングを行った結果、熱帯雨林地域で生じる森林火災のリスク評価は、PCAよりもSOMの方がより適切に表現（可視化）できることを明らかにした。特に、SOMでは数種類に分類されたSOMノードを用いて森林火災が生じるリスク分布の相違を確認することができた。しかしながら、SOMノードだけでは分類境界を明確に示すことができない場合もあり、SOMノードの彩色変化にDAを組み合わせて、SOMによる分類をより明確にすることを試みた。その結果、DAを用いることでSOM境界をかなり明確に表現することができた。したがって、SOMとDAを組み合わせた技法は、データを可視化してクラスタリングできる非常に有効な技法であり、熱帯雨林地域における森林火災が生じるリスクの分布を的確に表現できること

が分かった。

2. 周期的に実測された熱帯雨林地域の降雨に含まれる化学物質の濃度を評価するために、従来の技法であるAMMIモデルとSOMを降雨データの2つの分析要素(要素1:降雨継続時間・時期、要素2:化学物質の内容)の相互作用を見つけるために用いた。AMMIモデルは加法型と乗法型の相互作用効果を分散分析に組み合わせており、その分析結果は、分析要素の相互作用効果がF検定(有意水準5%)で有意となり、相互作用効果の相違を示すことができた。一方、この分析技法では要素2に関するAMMI散布図を使用することができないので、SOMを用いて降雨に含まれる化学物質の濃度を評価すると、降雨継続時間・時期に関係する降雨中の化学物質に関する相互作用のパターンの相違を明らかにすることができた。

3. 熱帯気候に特有な変動は、熱帯雨林地域において不定期的に生じる降雨パターンの変化に影響を及ぼすことから日降雨変動を適切に予測するための解析技法を開発することは極めて難しい。ここでは、熱帯雨林地域における日降雨量を的確に推定するために、人工ニューラルネットワークとファジィ推論を組み合わせたニューロファジィ分類モデルを構築して、熱帯雨林地域における日降雨パターンの推定を試みた。学習後のより簡潔なルールを用いて日降雨量における変動パターンを高精度で推定できることを明らかにした。このように複雑なデータに比較的簡単なニューロファジィ分類モデルを適用するだけで、ある程度データマイニングが可能となることを示した。

4. 一般に、熱帯雨林地域における気候と降雨の非線形性は、熱帯雨林地域の降雨パターンの複雑さに反映するが、雨季の降雨量の変動をモデル化するためにニューロファジィ推論システムを構築した。まず、降雨量の変動について、前述のファジィ推論モデルを用いて、よりの確な考察を可能にする簡単なファジィ分類ルールを作成した。このファジィルールに基づいて、降雨量の変動を予測するために適応型ニューロファジィ推論システムを構築し、平均2乗平方根誤差(RMSE)を用いて予測モデルの精度を評価した。その結果、本予測モデルを使用して熱帯雨林地域における雨季の降雨量の年変動を予測する場合、高精度で推定できることが分かった。結果的に、提案した予測モデルの妥当性も検証されたといえる。

以上のように、本研究で得られた知見は、インドネシア等における熱帯雨林地域の森林火災のリスク評価、降雨中の化学物質の変動、複雑な降雨パターンについての確な予測ならびに考察に資するとともに、環境情報のデータデータマイニングに新たな技法を提示し、地域環境科学や環境・農業情報科学の分野の発展に寄与するものであり、最終試験の結果とあわせて、博士(学術)の学位を授与することを適当と認める。