7K-02

R 言語による中国の地域別における大気汚染物質の特徴分析

張 翔† 史 一華‡ 徐 海燕†

（福岡工業大学大学院 情報工学専攻）

1 はじめに

急速な経済成長によって、中国では大気汚染問題が深刻さを増してきており、各種の研究分析を必要としてきている[2)]。本稿では、発表されている中国の 188 地域の 2014 年 1 月からの PM10、PM2.5、オゾン O3 をはじめとする 6 種類の汚染物質[1)]に対する分析を行い、その結果を報告する。具体的に、月別に各種の汚染物質のデータをクロス集計した上で、主成分分析と k-means によるクラスタリングを行う。その結果に対して、決定木を用いて分類ルール抽出と、時系列図やランキング地図による可視化を行う。それらによって、地域別・季節別の特徴を明確にすることを目指す。

2 基本的事項

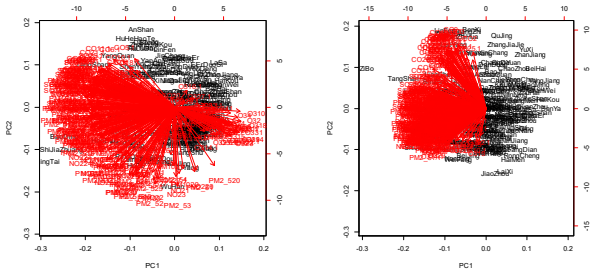
中国の 188 地域の 2014 年 1 月から 2015 年 7 月まで、季節別で PM10, PM2.5, SO2, CO, NO2, O3 の 6 種類の大気汚染物質及び空気質指数 AQI の値[1)]を、MySQL データベースにドウンロードし、MySQL 上のデータベースのデータを R 言語から取り出し、特徴分析を行う。データベースには、それらの都市の経度緯度などに関する情報も備えた。

3 大気汚染物質の特徴分析

3.1 主成分分析とクラスタリング分析

寒い季節と暑い季節の地域別の特徴を掴むために、まず、188 地域 2014 年の 1 月～2 月の計 59 日の 6 種類の汚染物質のデータを日別にクロス集計を行う。得られた計 354 個の集計データに対して主成分分析を行った結果が、図 1 (a) に示されている。O3 と CO, SO2 は完全に逆方向にある結果である。暑い季節 6 月～7 月の 61 日の計 366 個集計データに対する主成分分析の結果が図 1(b) に示されており、CO と O3 は同じく逆方向にあるが、SO2 がその両者と垂直に近い方向へ移動されていることが分かる。

さらに、その 2 つの季節のデータに対して、k-means によるクラスタリングを行い、図 2 に示されている結果が得られている。各クラスタに分類された都市の数は、表 1 に示されている。

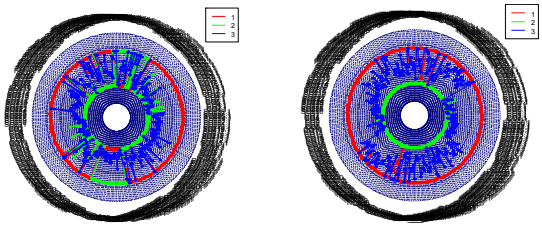


(a) (b)

図 1 2014 年二つの季節の主成分分析の結果

寒い季節では、赤色の外枠で表されている O3 以外の項目が高い種類の都市（下位）と緑色の内枠で表されている O3 だけが高い種類の都市(上位)と、ブルーで表されているその両者間にある種類の都市(中位)という 3 種類のクラスタになる(図 2(a))。

暑い季節では、O3 による違いがなくなり、全部が高い種類の都市（下位）と全部が高くない種類の都市（上位）と、それらの間にある都市(中位)という 3 種類のクラスタになる(図 2(b))。季節によって、3 種類のクラスタの性質が異なっていることが伺える。



(a) 1, 2 月 (b) 6, 7 月

図 2 2014 年二つの季節のクラスタリング 結果

表 1 各クラスタにある都市の数

時期/都市種類	下位	中位	上位
14 年 1 月～2 月	53 個	88 個	47 個
14 年 6 月～7 月	38 個	92 個	58 個

各クラスタにある都市の数については、表 1 に示しているように、寒い季節より暑い季節の方が、AQI が高い下位レベルに分類される都市の数が減り、AQI が低い上位レベルにある都市の数が増えている。3 種類の都市と地理分布状況の関係を掴

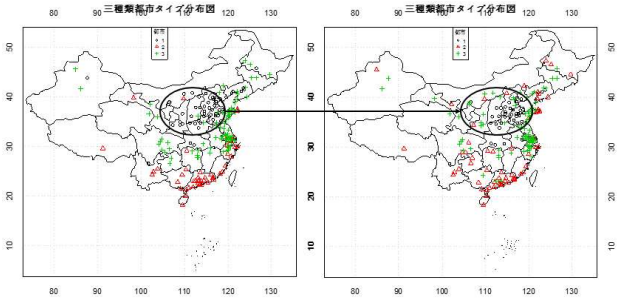
Analyzing Air Pollutants in Different Areas of China by R language

† Fukuoka Institute of Technology

‡ Seinan Gakuin University

Copyright ©2016 Information Processing Society of Japan.
All Rights Reserved.

むために、地図による可視化を行った。赤、緑、黒で示された上位、中位、下位に分類された3種類の都市の種類変化は、図3に示しているように、暑い季節になると、赤と緑色で示された上位、中位の都市が増えてくる。下位レベルから上位レベルに変わるという二段階アップという変化も起きている。



(a)1,2月　　(b)6,7月

図 3　2104年季節間で都市変換地図

3.2　決定木

クラスタリングの分類結果の原因を掴むために、クラスタリングの結果を目的変数とする決定木による分析も行った。寒い季節の都市分類は、次のような分類条件によって行われている結果が得られている。

- 上位レベルの都市
  SO2＜67＆O3≧83
- 中位レベルの都市
  SO2≧67&CO＜1.6　または
  SO2＜67＆O3＜83＆CO＜2.1
- 下位レベルの都市
  SO2≧67&CO≧1.6　または
  SO2＜67＆O3＜83＆CO≧2.1

すなわち、寒い季節ではSO2の値によって上位レベルの都市と中位、下位レベルの都市が分類され、COとO3の値によって中位と下位レベルの都市が分類されていることになる。一方、暑い季節においては、PM10、SO2、CO、NO2の値によって分類されている結果になっている。SO2は石炭の燃焼、NO2は車の排気ガスに密接に関係していることから、分類ルールが季節と地域特性を反映していることが伺える。

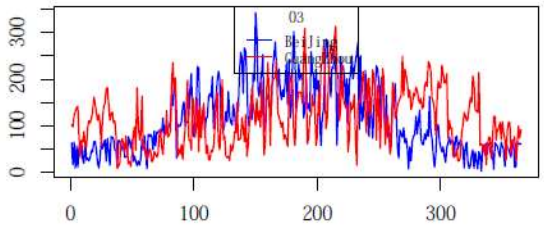


図 4　2014年北京と広州03の時系列図

3.3　時系列とランキング地図による可視化

3.1節のクラスタリングの分類結果をもう一歩確認するために、2014年1月から12月までの一部分の都市のAQIと6種類汚染物質の濃度の日平均値データの時系列による可視化を行っている。北京市と南に位置する広州市の2014年一年間の03の時系列図(図4)に示されているように、北京市は暑くなると03の濃度が増えるが、広州市は季節による変化は少ない。この性質が図2(a)の03が他の項目と逆になっている原因と思われる。

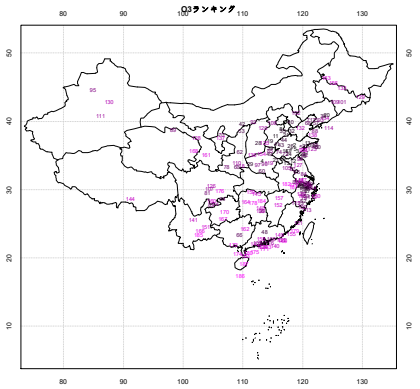


図 5　2014年6月～7月の03ランキング地図

188個の都市における各種の汚染物質のランキングを確認するために、地図による季節ごとのランキングの可視化も行っている。順位が前の方の都市は濃い色で表示している03のランキング図(図5)より、北に位置する都市の03ランキング順位が、寒い季節は暑い季節より上げている結果が確認できている。一方、03以外の5種類汚染物質のランキングは、暑い季節になると、一部の北に位置する都市の順位が下がり、一部分の沿海の南に位置する都市の順位が上げっていることが確認できた。

4 まとめ

本稿では、中国の188地域の2014年1月からの6種類の大気汚染物質に対する主成分分析、クラスタリングによる分析を行った。その結果を決定木による分類ルールの抽出と、時系列やランキング地図による可視化を行った。冬と夏の季節別・北と南の地域別における大気汚染物の特徴をある程度掴むことができた。春と秋も含めた四季における分析などは今後の課題である。

参考文献

1) http://aqistudy.sinaapp.com/historydata/daydata.php
2)王雅潞、曾怡、余加奎:基于大气环境区域性污染特征分析对“联防联控”体制实施的建议 中国統計学会 2013

Copyright ©2016 Information Processing Society of Japan.
All Rights Reserved.