

自主課題研究概要

～ 3 値のデータのみを用いた天気予測～

電子情報学類 3 年 2 4 5 林 直樹

1. 研究課題

Rを用いて、サポートベクターマシンで分類を行うことに寄り、天気のパred測を行う。

2. 研究に用いたデータ

Yahoo!天気情報(<http://weather.yahoo.co.jp/>)からダウンロードした北陸6都市の2004年後半～2008年までの天気データを利用した。また、雪の日も雨とした。

3. 研究内容・方法

以前の自主課題研究で、各地の天気・気象情報を特徴ベクトルの数値データに加工、晴れ～曇り～雨をそれぞれ 1～0～-1 に変換し、サポートベクター回帰を行って天気を予測していたが、天気データだけで分類によって天気を予測してみようと考えた。

Rについては kernlab パッケージの ksvm 関数を使用。カーネルはデフォルトの RBFカーネル。精度の評価は leave-one-out class-validation で行った。

3. 結果・考察

データ全体での予測の正解率は 75.4%だった。また、都市別や月別についての正解率も求めた (図 1～3 に示す)。

全体を用いて行った分類では正解率が 75.4%とそれなりに予測を当てることがで

きた。

都市別や月別の正解率をみると、曇りと雨の正解率が不安定だった。また、どちらも総合的な正解率はあまり曇りや雨の正解率に影響を受けずに安定しているが、晴れの日が圧倒的に多いためほとんど晴れの正解率に依存していると考えられる。そこで、月別に行った分類で正解率に大きな差がでたので、各月の天気の比率を見たところ、やはり晴れる日の多い月は正解率が高く、逆に少ない月は正解率が低いようだった。

4. まとめ・感想

全体で 75%、都市別にみても 7 割以上とそれなりに安定して正解していると考える。

北陸においても晴れの日が圧倒的に多いというのは少し意外だった。しかし、予測がほとんど晴れに偏ってしまい、曇りや雨の予測の精度が低くなってしまったのは残念だった。

R を用いることで簡単に分類による予測ができるのは便利だった。

反省としては、曇りや雨などの精度を上げる方法や、都市によって雨や曇りの正解率に大きな差がある原因についても調査できればよかった。

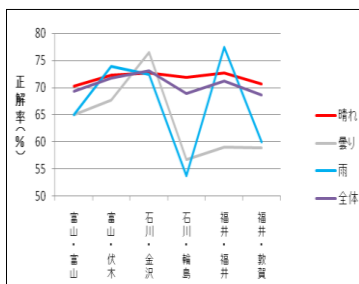


図 1:都市別の正解率

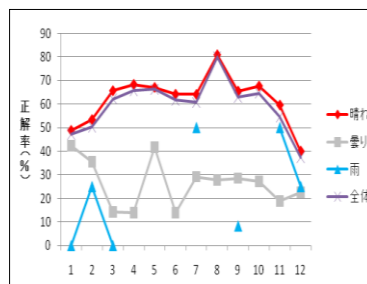


図 2:月別の正解率

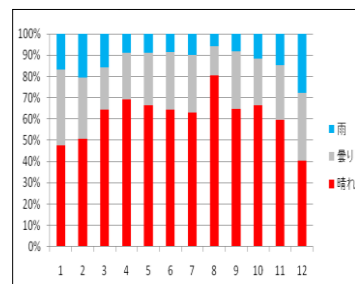


図 3:各月の天気の比率