



2008 年度上期未踏 IT 人材発掘・育成事業(未踏ユース)採択案件評価書

1. 担当PM

安村 通晃 PM(慶應義塾大学 環境情報学部 教授)

2. 採択者氏名

チーフクリエイター: 木村 広希(筑波大学大学院システム情報工学研究科
博士前期課程)

コクリエイター: なし

3. プロジェクト管理組織

株式会社 創夢

4. 委託金支払額

2,999,525 円

5. テーマ名

気圧配置の自動分類・検索システムの開発

6. 関連Webサイト

なし

7. テーマ概要

テレビや新聞の天気予報において、気圧配置というものをよく目にする。この気圧配置とは高気圧と低気圧の分布状況を意味するものである。「冬型の気圧配置が強まると、強い風が吹き込み非常に冷え込む」という気象現象は有名なものの一つである。

ここで、このような気象現象についての研究に着目してみると、研究者たちには悩ましい問題がある。

研究者が研究を行う際に、ある特定の気圧配置の事例を多数必要とすることがあり、このとき彼らは、数十年分の過去のデータの中から必要なデータを探す。これはつまり、各データがそれぞれ必要とする気圧配置であるかを一枚一枚判別し、分類することを意味する。この気圧配置の分類は、私の知る限り、現状では全ての研究者が目視による判別で行っている。そのため、必要とするデータの数が多いため、この目視による判別という作業は過酷なものとなる。

そこで、この作業の負担を軽減すべく、気圧配置を自動で分類し、検索できるシステムを開発する。

8. 採択理由

気象予報ほどコンピュータの計算力が発揮できるものは無いだろうし、実際、数値計算や高速演算の対象としては気象予報は典型的な応用の一つであった。ところが、現実の天気図が西高東低とか梅雨型である、などという分類は人手で行われてきたようだ。

これを、コンピュータを用いて自動分類し、気象学の研究に活用しようとするのが、本開発の提案内容である。これは、まさにコロンブスの卵、ともいべきプログラム開発の提案であって、なんでこのようなことに今まで気付かなかったのであろう、という思いが強い。

これは、分類問題なので、SVM(サポートベクターマシン)の方法が効果的なアプローチである。しかも、これまで、1981年から2000年までの分類済みのデータを使うこともできるという。いったんコンピュータで分類した結果で、曖昧性が残るものに関しては、人手で分類して、その分類精度を向上させるという、確実な方法を採用する予定である。さらに、個人毎の分類結果も登録し、個人差を比較するような内容まで含んでいる。気象関係に詳しい研究者の協力も得られるそうで、開発の成果が大いに期待できる。

9. 開発目標

本システム開発のため、本プロジェクトでは以下の実装・開発を行なう。

1. 各気圧配置の特徴抽出手法の検討
2. 分類実験による最適な分類器の作成
3. 検索性 Web ページの構想
4. 検索性 Web ページの実装

10. 進捗概要

当初、SVM(サポートベクターマシン)により、気圧配置を分類することは極めて容易と考えていたが、実は、利用する気象データは、気圧、風速、温度などの情報のみであり、そもそも低気圧の位置や高気圧の位置を見つけたり、台風や前線なども判定しなければならない。そのような事情で、本プロジェクトは、開始から間もなくして、本来の気圧配置分類までにはなかなか大変で、実際、プロジェクトレビューも、現地レビューとPM側に来てもらうのを合わせて、二度行なった。最終的には、6種類の気圧配置の分類はほぼ終わることができたし、個人別の判定システムのベースの部分も仕上げることができた。

11. 成果

気圧を中心とする気象データ(JRA-25と呼ぶベクトルデータを生成)から、西高東低冬型、低気圧型、移動性高気圧型、停滞前線型、南高北低夏型、台風型の6種類の気圧配置を判別する。この場合、与えられた気象データが、この6種類のいずれかになるかという分類問題ではなく、それぞれの気圧配置に属するかどうかの2クラス分類である。ここで取り扱う6種類の気圧配置の典型例を図1～図6に示す。

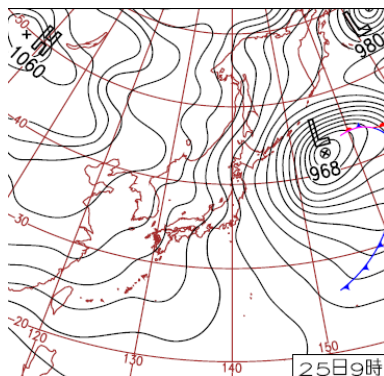


図1 西高東低冬型

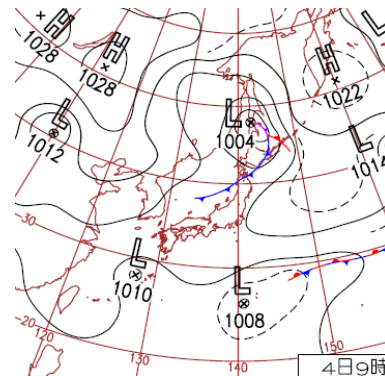


図2 低気圧型

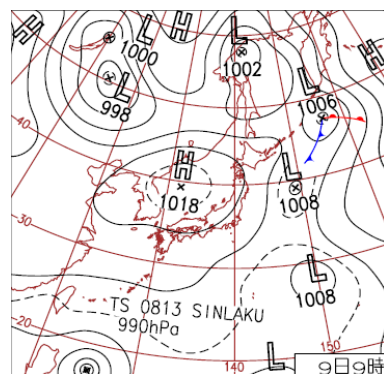


図3 移動性高気圧型

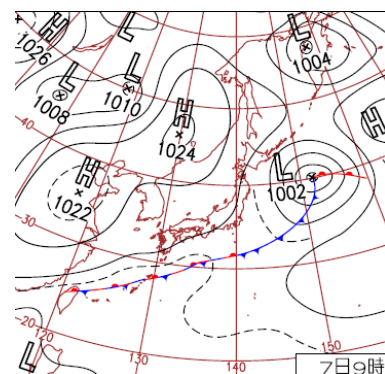


図4 停滞前線型

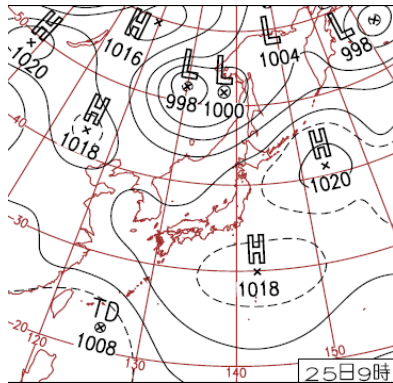


図5 南高北低夏型

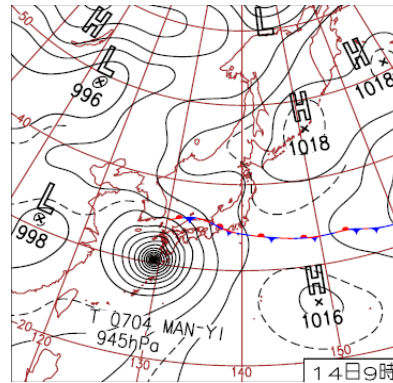


図6 台風型

学習データ(クラスラベルデータ)としては、1991年～2000年の間の人手による気圧配置分類データを用いた。テストデータとしては、1981～1990年の気象データを用いた。ベクトルデータとクラスラベルデータにより、TinySVMを用いて分類を行なった。その結果の評価として、次の3つの指標を用いる。

適合率 = $TP / (TP + FP)$ 、再現率 = $TP / (TP + FN)$ 、F値 = $2TP / ((TP + FP) + (TP + FN))$

ここで、TPとは、例えば正解が冬型の場合結果も冬型になる場合。FPは、結果は冬型だが正解が非冬型の場合。FNは、正解が冬型なのに間違っ非冬型の結果になる場合。TNは、結果も正解も非冬型の場合。

SVMで分類させた結果の指標は、現状では表1の通りとなった。

表1 SVMでの分類結果

気圧配置	適合率	再現率	F値
西高東低冬型	0.89	0.83	0.86
移動性高気圧型	0.81	0.73	0.76
低気圧型	0.77	0.63	0.69
停滞前線型	0.63	0.76	0.69
南高北低夏型	0.70	0.49	0.58
台風型	0.80	0.27	0.41

この結果を用いて分類した気圧配置に対して、Web上で検索できるシステムを構築した。その画面例を図7に示す。

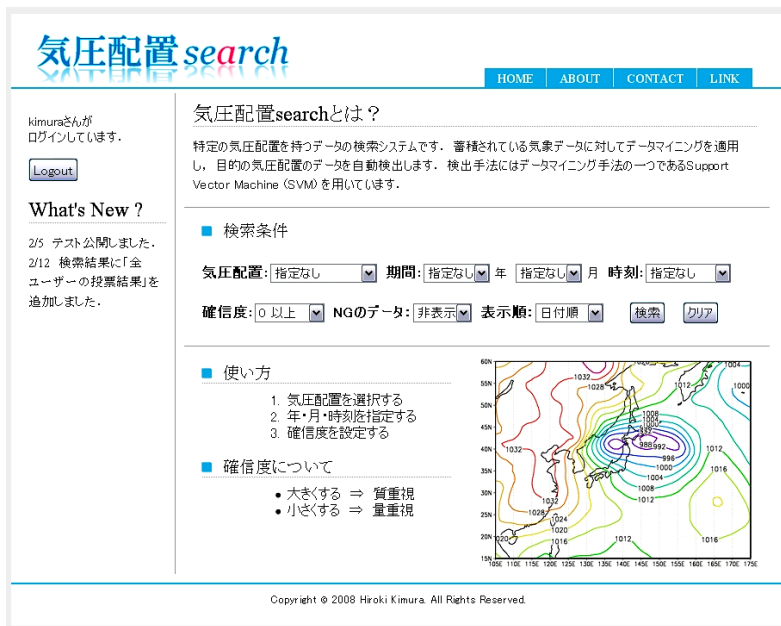


図 7 気圧配置検索 Web ページの画面例

この検索システムは、熟練したユーザーが利用するログイン状態と、初心者ユーザーが用いる非ログイン状態がある。いずれの場合も、気圧配置、期間、時刻、(システムの)確信度などを入力して検索する。ログイン状態では検索結果が望ましいか否かを投票することができるが、非ログイン状態ではできない。図 8 に検索結果の一例を示す。



図 8 気圧配置検索結果の例

この場合、該当する気圧配置データが 17 件見つかった。投票状態も示されている。また、前後の時刻における気圧配置画像も可視化して表示することができる(図 9)。

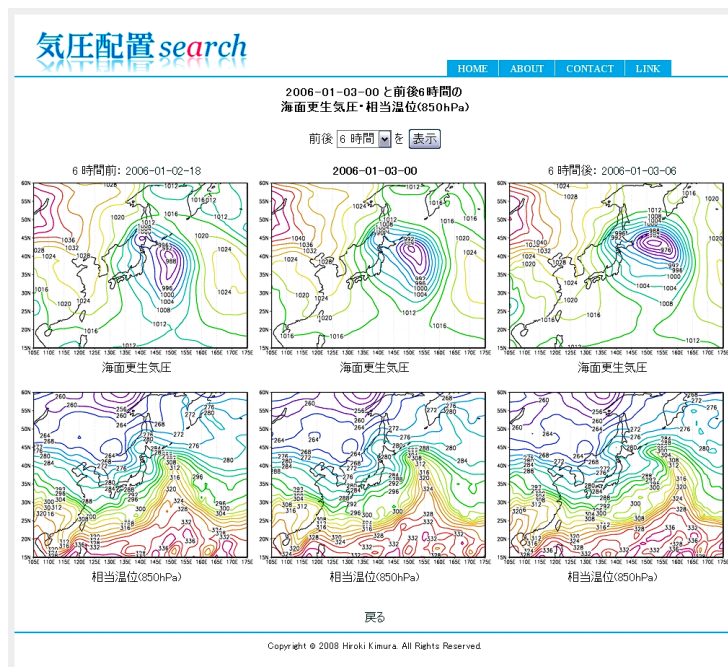


図 9 前後の時刻の可視画像表示ページ

さらに、今回作成した気圧配置検索システムを、気象を専門に学ぶ学生 23 名に使ってもらった評価を行なった。その結果、気圧配置毎の満足度は図 10 に示す通りとなった。

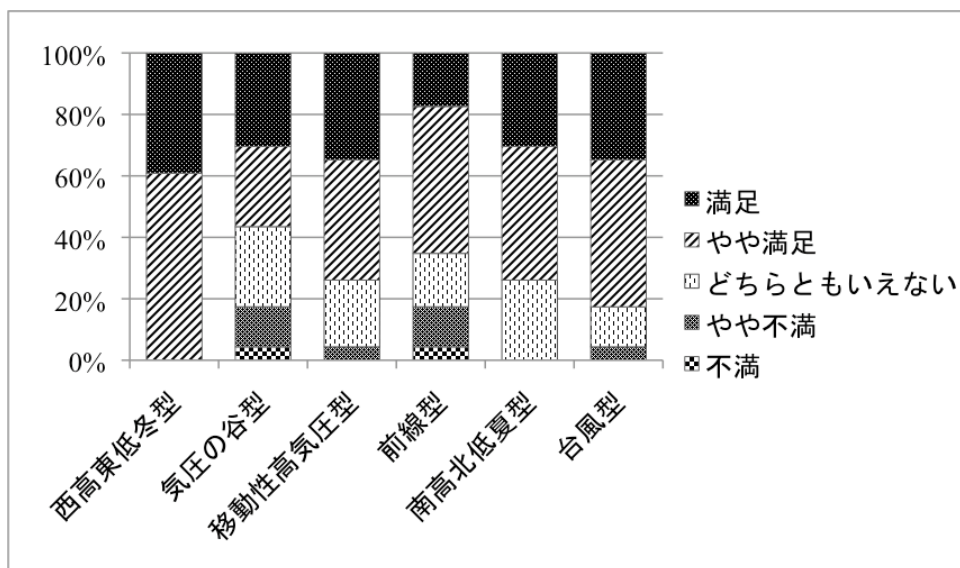


図 10 気圧配置毎の検索結果満足度

さらに、システムの使いやすさについては満足 31%、やや満足 61%を合わせると 92%の利用者がほぼ満足している。特定の気圧配置の事例収集は有益だと思うかに関しては 78%がそう思うと答えている。この検索システムを利用したいかに関しては、57%がそう思うと答えたが、どちらとも言えないも 43%を占めている。

12. プロジェクト評価

気圧配置の自動分類など、とうの昔に実現していたかと思っていたが、意外や意外。まだ人手で行なっていたそうである。今回は、人手による分類をコンピュータによる自動分類のさきがけとなる研究開発である。このシステムを用いると目視(4 時間)に比べて、少なくとも半分の 2 時間で済むことが分かった。精度に関しては今後改善していく必要があるが、コンピュータによる自動分類と、人手での判定の補助という大きな目標の第一歩となる研究開発が見事に達成された。この開発プロジェクトの意義は大きい。

13. 今後の課題

気圧配置によっては、すでにかかなりの分類精度を達成しているものと、そうでないものがある。さらに、精度向上の工夫が必要であろう。元となる気象データのうち、気圧だけで良いのか、風速や気温のデータも用いた方が良いのか、時間的な変化に着目した分類精度の向上など、さらなる改善・改良が望まれる。