

ニューラルネットワークによる桜の開花日の予測に関する考察

A Study on the Prediction of Cherry Blossoms Bloom Date using Artificial Neural Networks

檀 裕也 川井 東
Yuya DAN Akira KAWAI松山大学 経営学部
Faculty of Business Administration, Matsuyama University

桜の開花は、過去の気温や日照時間、春先の気温上昇などの条件によって決まると考えられている。さまざまな経験則に基づく予測の計算式が提案されているものの、桜の開花日を正確に予測することは難しい。本研究では、桜の開花予測にニューラルネットワークを適用した結果を紹介し、その有効性と限界について述べる。

The day of blooming for cherry blossoms is considered to depend on past temperature, sunshine duration, and liquid precipitation etc. Although the law or formulas are proposed that are based on experiences for the prediction, it is difficult to predict correctly the day of blooming for cherry blossoms every year. This paper introduces the result of the prediction of cherry blossoms blooming using artificial neural networks with past temperature data, then describes the efficiency and the limit of the prediction.

1. はじめに

毎年、気象庁*1や民間の気象情報会社から桜の開花予測が発表されている。バラ科サクラ属のソメイヨシノが約8割を占める日本の桜は、全国各地の開花日や満開日などが予測されると、花見客にとっては3月から5月にかけての桜前線として北上する様子が春の到来を期待させる。開花日を正確に予測することは、天気予報と同様に、花見客の行動計画のみならず、各種イベントを企画する旅行業界や飲食業界などビジネスにおいて季節的需要に応えることができる。

しかし、ソメイヨシノの開花日を正確に予測することは、非常に難しい。生物学における植物の成長モデルとして、過去の気温や日照時間、降水量、春先の気温上昇などの条件によってソメイヨシノの開花が決まると考えられている。経験則に基づく予測の計算式や回帰分析による法則が提案されているものの、ソメイヨシノの開花日を正確に予測するという意味で定量的な法則が定式化されるまでには至っていない。

本研究では、このような状況にあるソメイヨシノの開花予測において、非線形現象の解析に有効なニューラルネットワークを適用した結果を紹介し、その有効性と限界について述べる。

2. 問題の背景

松山におけるソメイヨシノの開花日は、表1のように推移している。ここで、開花日は3月1日を基準とする整数値を用いている。したがって、3月中の開花は開花日と日数が一致するが、開花日が32のときは4月1日などと読み替える。

ソメイヨシノの開花予測に関する先行研究によると、植物としての生育に依存する要因として春先の気温上昇が大きく関与していると考えられている。実際、過去の気温や日照時間、降水量などの条件について開花日との関係を分析した研究[川井]によると、気温、特に開花前にあたる春先の気温上昇が大きく影響していることが示唆された。

そこで、予備調査として、松山地方気象台におけるソメイヨ

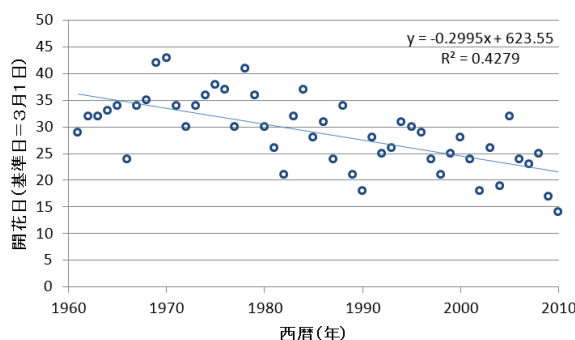


図1: 松山における1961年~2010年のソメイヨシノ開花日【松山地方気象台】

シノの開花日について、松山における1月、2月および3月の平均気温を説明変数とする回帰分析を行ったところ、3月の平均気温を説明変数とするときが最も高い決定係数 ($R^2 = 0.54$) を得た。

3. 多層パーセプトロン

ニューラルネットワークは、動物の脳を数学的にモデル化した機械学習の一つで、多数のニューロンが結合したネットワーク構造によって問題を解決する手法である。線形モデルに対応する単層パーセプトロンは、線形分離可能な問題に対して有効性を示したが、分類や回帰分析では解決できない問題が多く存在していた。入力層と出力層の間に、隠れ層を挟んだ多層パーセプトロンでは、線形のみならず任意の関数を表現できることが知られ、単層パーセプトロンでは解決できなかったいくつかの現実的な問題を解決し、1986年以降、実際の問題解決に対する有効性が示された。[Rumelhart]

多層パーセプトロンの典型例として、入力層 ($i = 1$)、隠れ層 ($i = 2$) および出力層 ($i = 3$) の3層からなる階層型ニューラルネットワークについて考察する。いま、第 i 層 ($i = 2, 3$)

連絡先: 檀 裕也, 松山大学 経営学部, 愛媛県松山市文京町 4-2, 電話:089-925-7111 (代表), dan@cc.matsuyama-u.ac.jp

*1 気象庁は、正確な予測ができないとして、2010年を最後に開花予測から撤退した。

表 1: 実行環境

CPU	Intel Core 2 Duo T9600 (2.80GHz)
メモリ	4.00GB
OS	Microsoft Windows Vista (32 ビット)

のユニット j における信号がシグモイド関数

$$y_{ij} = \frac{1}{1 + \exp(-x_{ij} - \beta_{ij})} \quad (1)$$

によって伝達されると仮定する。ただし、 β_{ij} はユニット ij の閾値である。すると、ユニット内部における信号の伝達は、区間 $(-\infty, \infty)$ から区間 $(0, 1)$ への写像としてとらえることができる。なお、入力層 ($i = 1$) には入力データが与えられる。

第 i 層 ($i = 2, 3$) のユニットは第 $i - 1$ 層に属するすべてのユニットから出力された信号を受け取る。このとき、ユニット j とユニット k の結合の強さを表す重み w_{ijk} を使って

$$x_{ij} = \sum_{k=1}^{N_{i-1}} w_{ijk} y_{(i-1)k} \quad (2)$$

のように変換された信号が次のユニットの入力信号となる。

逆伝播法による学習では、教師信号として入力データごとに正解 z_j ($j = 1, \dots, N$) を与え、ニューラルネットワークが出力する信号との誤差

$$e_j = \frac{1}{2} (y_{3j} - z_j)^2 \quad (3)$$

を評価し、最小二乗法によって誤差の総和

$$\sum_{j=1}^N e_j \quad (4)$$

が最も小さくなるように重み w_{ijk} を最適化する。その際、重み w_{ijk} の値を変えながら、ニューラルネットワークが出力する信号と入力データが近づくように繰り返し学習する。

4. 実験

本実験は、表 1 に示した計算環境において実行した。

入力に用いたデータは、松山における 1980 年から 2009 までの過去 30 年間における 3 月の日別平均気温である。また、松山におけるソメイヨシノの開花日 (基準日 = 3 月 1 日) を教師信号 (正解データ) として与えた。したがって、ニューラルネットワークは 31 次元ベクトルの入力に対し、スカラー値を出力することになる。なお、隠れ層は 64 ユニットで構成した。

統計解析システム R バージョン 2.7.2 の neural パッケージを利用し、mlptrain 関数を用いて学習し、mlp 関数を用いて過去 30 年分の開花日を検証するとともに、2010 年における 3 月の平均気温データから開花日を予測する。

4.1 学習

毎年 3 月の平均気温データは、31 日分のベクトルとして読み込み、全データの最大値と最小値を使って 0 以上 1 以下の値にスケール変換した。また、開花日は 3 月 1 日を基準とするスカラー値 (整数) として、平均気温データと同様に、開花

日の最大値と最小値を使って 0 以上 1 以下の値にスケール変換した。平均気温データを入力層、開花日を出力層とし、隠れ層 (1 層) のニューロンを 64 ユニット分用意することによって学習させた。

```
>mlptrain(入力層, neurons=64, 出力層, it=10000)
```

なお、学習回数は 10 から 10000 まで変化させながらニューラルネットワークの学習を繰り返し、学習回数による精度の向上について調べるとともに、適用の妥当性を検証する。

4.2 適用

mlptrain 関数を使って過去 30 年間の気温と開花日の関係を学習させたニューラルネットワークにおいて、入力データを使ってソメイヨシノの開花日について学習の精度を検証する。また、2010 年 3 月の気温データを入力として与え、開花日を予測する。そのとき、mlp 関数に学習で得られた重みを使い、伝達関数にはシグモイド関数 (1) を指定した。

```
>mlp(入力層, weight, distortion, neurons, actfns)
```

それぞれの予測値が実際の開花日をどの程度の精度で当てているのかを調べた。

5. 結果

学習の精度を評価するため、入力データを与えたときのニューラルネットワークによる出力と教師信号の誤差を表 2 にまとめた。西暦ごとに、実際の開花日、ニューラルネットの適用結果、および両者の誤差 (適用結果 - 実際の開花日) を示している。

実際の開花日とニューラルネットワークによる出力は、学習の繰り返し回数を増やすことで誤差が小さくなる。表 3 は、学習回数ごとに学習精度と開花予測日を表している。学習精度 σ とは、ニューラルネットワークによる学習の結果として、実際の開花日 D_j と適用結果 d_j がどの程度近いのかを表す指標

$$\sigma = \sqrt{\frac{1}{N} \sum_{j=1}^N (d_j - D_j)^2} \quad (5)$$

である。入力データを正確に再現しているときには $\sigma = 0$ となり、ずれが大きくなると入力データの個数に関係なく σ は大きな値を取る。

図 2 は、横軸に学習回数 (繰り返し) の対数、縦軸に学習精度を取ったグラフである。実験の結果、学習回数 (繰り返し) を増やすごとに学習精度が高まっていることが分かった。

一方、2010 年の平均気温データにニューラルネットワークを適用した開花予測日は、実際の開花日から大きく乖離していることが分かる。学習回数 (繰り返し) に関係なく開花予測日は収束するようには見られないことから、未知のデータに対する予測が正確でないことが示唆される。

6. まとめ

本稿では、松山における 3 月各日の平均気温ベクトルを入力値、ソメイヨシノの開花日を出力値とする教師信号を与えた場合のニューラルネットワークについて議論し、学習回数 (繰り返し) を増やすことによって学習の精度は高まるものの、未知の平均気温データに対する予測が十分でないという知見が得られた。

表 2: 学習精度の評価結果 (N=200)

西暦	開花日	適用結果	誤差
1980	30	31.3	1.3
1981	26	27.8	1.8
1982	21	20.0	-1.0
1983	32	30.9	-1.1
1984	37	34.1	-2.9
1985	28	28.6	0.6
1986	31	32.1	1.1
1987	24	23.2	-0.8
1988	34	31.4	-2.6
1989	21	22.7	1.7
1990	18	21.1	3.1
1991	28	24.7	-3.3
1992	25	23.2	-1.8
1993	26	27.4	1.4
1994	31	30.6	-0.4
1995	30	26.4	-3.6
1996	29	28.2	-0.8
1997	24	24.3	0.3
1998	21	22.6	1.6
1999	25	21.8	-3.2
2000	28	25.8	-2.2
2001	24	23.6	-0.4
2002	18	20.2	2.2
2003	26	28.6	2.6
2004	19	20.5	1.5
2005	32	30.3	-1.7
2006	24	26.6	2.6
2007	23	22.3	-0.7
2008	25	23.8	-1.2
2009	17	20.0	3.0

表 3: 学習精度と 2010 年気温データへの適用結果

学習回数	学習精度	開花予測日
10	4.74	25.27
20	4.45	23.15
50	3.30	26.56
100	2.44	25.20
200	2.00	27.59
500	1.64	29.29
1000	1.13	31.10
2000	0.81	29.42
10000	0.10	33.49

学習精度

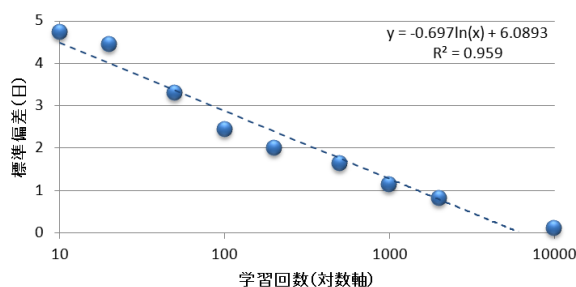


図 2: 学習の繰り返しによる学習精度の推移

2010 年の松山におけるソメイヨシノの開花日は 3 月 14 日と例年に比べて早いという事実を踏まえると、入力データで与えた 3 月の平均気温ベクトルの成分のうち半分も反映されていないことから、予測の精度が低く留まった可能性が指摘される。開花日予測、すなわち、開花日が確定されていない段階でのニューラルネットワークの適用には、開花直前までの気温の上昇を反映させる必要があるため、1 月～3 月といった長期の平均気温データを入力値として与え、ニューラルネットワークの重みづけによって柔軟に対応させることが一つの改善策であると考えられる。すると、相関係数の大きな説明変数を自動的に選び出し、必要に応じて計算に取り入れたり取り入れなかったりする機構が実現できる可能性がある。

今回の研究は、松山における気温とソメイヨシノの開花日の関係について議論したに過ぎないが、松山以外の全国各地のデータを用いることによって、統計的にニューラルネットワークによる開花日の予測について評価することが今後の課題として求められる。また、学習だけでなく、ニューラルネットワークによる法則の発見と表現に関する知見を深めたい。

参考文献

- [Rumelhart] D. E. Rumelhart, G. E. Hinton, R. J. Williams, Learning internal representations by error propagation, Parallel Distributed Processing, Vol. 1, pp. 318-362. (1986)
- [川井] 川井東, 桜の開花の分析, 卒業論文. (2010)
- [川上] 川上皓史・山尾僚・森岡耕一・池田博・波田善夫, 温度変換日数法を用いたソメイヨシノ(バラ科サクラ属)の開花調節, Naturalistae 13: pp. 1-7. (2009)
- [豊田] 豊田秀樹, データマイニング入門, 東京図書. (2008)
- [中野] 中野良平, ニューラル情報処理の基礎数理, 数理工学社. (2005)
- [R] R, <http://www.r-project.org/>
- [気象庁] 気象庁, <http://www.jma.go.jp/>

