

# 台風経路の文字列化による類似度順位付け

木村佳羽（指導教員：神山翼）

## 1 はじめに

台風経路の類似度を調べ過去の事例を参照することは、自然現象についての理解を深めるだけでなく、防災上の観点からも重要である。クラスタリングの手法で台風経路をグループ分けする研究 [1] や、DTW 法を用いて二つの台風経路間の距離を定義した研究 [2] などが既に提案されているが、今後も増え続ける台風経路データを効果的に処理するには、簡便でより扱いやすい手法を模索することも有意義であろう。

本研究では、台風経路についての自由度の高い検索や分類を少ない計算量で実現させるために、時系列データを文字列で表現する二次元の SAX 法 [3] を用いた。日本に接近する台風経路をその中心位置の緯度と経度について文字列に変換し、その文字列の比較に基づいて台風経路に類似度の順位付けを行う方法を提案する。

## 2 提案手法

### 2.1 台風経路の文字列化

気象庁より公開されているベストトラックデータ [4] の 1951 年から 2019 年までの 69 年間の全ての台風について、中心位置の緯度と経度の値を読み取り、以下の手順で台風経路の文字列化を行った。

#### 2.1.1 領域とその分割

台風経路を文字列化するには、台風が通過しうる日本付近（東経 80°～西経 160°、北緯 80°～南緯 10°）の領域を図 1 のように 10° 間隔の等緯度等経度格子に分割し、その各格子に文字を割り当てておく必要がある。アルファベットなどを割り当てるとは格子数に対して文字数が足りなくなるため、十分な文字数の漢字のリストを用意し、右上の格子から順に割り当てた。

格子間隔を大きくするほど類似度の判定は甘くなるが、小さくしすぎても、一つの格子内に入るデータ数が少なくなり類似台風が抽出できない。5°、10°、15° の 3 通りを試し、今回対象とする領域に最適な格子間隔として 10° を選んだ。

#### 2.1.2 台風経路の文字列化

ベストトラックデータは 6 時間ごとに解析されているため、6 時間ごとの台風の中心位置がどの格子内であるかによって格子に対応する文字を順に並べ、台風経路を文字列に変換した。例えば、2019 年 10 月に甚大な被害を出した台風 19 号（令和元年東日本台風）の経路は「寸寸寸水水水水水元元元元元元元元元元牛切切切切切切切切切切心心心心心心心引引引六六六犬久千千下」という文字列に変換される。

### 2.2 経路類似度の算出

文字列化した台風経路の類似度を定量化するために、二つの文字列の類似度を計算する手法の一つであるレーベンシュタイン距離を用いる。レーベンシュタイン距離は、1 文字の挿入・削除・置換のいずれかを 1 手順として、一方の文字列をもう一方の文字列に変形

するのに必要な手順の最小回数として定義されている。二つの文字列が似ているほど距離は小さく、異なるほど距離は大きくなるが、そのままでは文字列の長さを考慮しない。

提案手法では、寿命の長い台風ほどその台風経路は長い文字列に変換される。台風ごとの寿命の差を考慮するために、二つの文字列のうち長い方の文字数でレーベンシュタイン距離を割り、正規化を行った。また、正規化した値を 1 から引いて 100 倍したものを経路類似度 (%) として扱う。

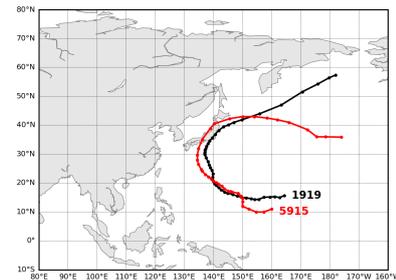


図 1：伊勢湾台風と令和元年東日本台風の台風経路図

例えば、図 1 に経路図を示した二つの台風（1959 年台風 15 号（伊勢湾台風）、令和元年東日本台風）の経路類似度は 60.42% と算出される。

## 3 類似台風

令和元年東日本台風に着目し、算出した経路類似度と軌跡類似度に基づく類似台風を定量的に抽出した。

### 3.1 格子の位置

ある 2 点間の距離が 9 度離れていても同じ格子内であれば両方に同じ文字が割り当てられるのに対し、たった 0.1 度離れただけでも、それが格子と格子の境界付近であれば異なる文字が与えられてしまうことがある。

提案手法では格子間隔を超える文字間の地図上の距離を考えないため、このような偏りが解消されるよう、格子を適当な位置に動かすことを考える。台風経路の類似部分が最も多く同じ格子内に含まれるような格子の位置を毎回探すことは計算量の増加につながるため、図 1 での格子の位置 (A) を緯度、経度、緯度・経度方向にそれぞれ 5° ずつずらした状態（これらを格子位置 B, C, D とする）で改めて台風経路の文字列化を行い、格子位置ごとの経路類似度を比較した。

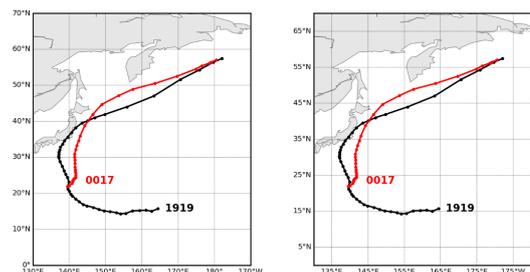


図 2：2000 年台風 17 号と令和元年東日本台風の台風経路図

