

機械学習を用いた パターン抽出及び気象予測



総合情報学科 知能情報システム研究室

永井ゼミ 4年

後藤遼平



動機

- 現在使用されている抽出及びその利用方法は、主に出現頻度や回数を用いるものが多い。
- 間隔や期間といった要素を用いるものは、回数を用いるものとは別の計算式や特定手法を用いるものが多く見つかった。
- 上記の時間軸のような要素を用いた抽出を回数を用いたものと同じ手法で行うことで、精度の高い推測、及びその推測に基づいた数字列などの生成がより容易に実現可能なのではないかと考えた。



目的

- 入力されたデータから特定のパターンや法則を探し、データの続きを予測、生成
- グラフなどで視覚的に表示



予測の概要

- 入力されたデータから、出現回数などを基に特定の値やその組み合わせを抽出
- 抽出したものの間隔や出現頻度から出現パターンを予測、提案
- 提案に従ってデータが続いた場合の出現位置を予測
- 予め分割しておいたデータや入力されたデータに適用するなどの方法で精度を評価



学習に使用するデータ

- 気象データの使用を検討
- 上記データの入手はデータカタログサイト「DATA GO JP」
(<https://www.data.go.jp/about-data-go-jp>) で行う



検証・評価方法

- 平均値を取るなどして、変化を一定にしたデータを入力
- 乱数を用いて作成したデータを入力
※前者は高評価、後者は低評価を予想
- 学習に使用していないデータを用意し、データの前半を入力
生成された予測を後半と比較して評価



先行研究について

- 先行研究では
 - ・ 株価をはじめとする市場予測
 - ・ 気象予測
 - ・ 燃料電池などの損傷予測などがみられた。
- 本研究は、これらの導入や時系列要素の利用に対して、理解や実装を容易にすることを目的とする。

研究計画

1	2	3	4	5	6	7	8	9	10	11	12	1	2	3
---	---	---	---	---	---	---	---	---	----	----	----	---	---	---

言語の学習

試作システム
作成

検証

論文作成

試作システムの使用言語・ツール

Python

Anaconda3



7月現在の状況

- 先行研究で使用されている数式やプログラムを勉強中
- 試作システムは、回数を用いたものは完成(低精度)
時系列データを用いたものは未完成



今後の予定

- 先行研究や類似プログラムの理解、同様のプログラムの作成
- 上記プログラムを参考に、データの加工や処理方法の決定
- 収集したデータが上記の加工、処理方法で利用できるかどうかの確認
- 試作プログラムの作成、精度などの改良