Robust and interpretable unsupervised machine learning techniques for analyzing the climate system

Derek DeSantis[†], Phil Wolfram, Boian Alexandrov, Raviteja Vangara, Erik Skau, Katrina Bennett

March 5, 2021

Introduction

Difficulties with Machine Learning - ML Safety

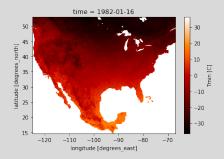


Figure: OpenAI CoastRunners misspecified reward function

ロト (個) (日) (日) (日) 日 のへの

Climate Biome Clustering

L15 Gridded Climate Dataset - Livneh et. al.



- Gridded climate data set of North America.
- Grid cell is monthly data from 1950-2013, six kilometers across.

ロ > < 同 > < 三 > < 三 > 三 - シへで

• Available variables used: precipitation, maximum temperature, minimum temperature.

- Climate Biome Clustering
 - Difficulties with Machine Learning ML Safety

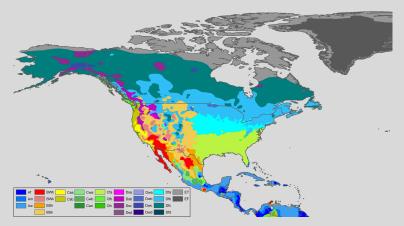


Figure: Köppen-Geiger map of North America (Peel et. al.)

ロン (得) (ヨ) (ヨ) (ヨ) (ヨ) の(の)

Climate Biome Clustering

└─Problems with Köppen-Geiger

Problem

• Climate depends on more than temperature and precipitation.

ロン (得) (ヨ) (ヨ) (ヨ) (ヨ) の(の)

Climate Biome Clustering

└─Problems with Köppen-Geiger

Problem

• Climate depends on more than temperature and precipitation.

ロン (得) (ヨ) (ヨ) (ヨ) (ヨ) の(の)

• Can only resolve land.

Climate Biome Clustering

└─Problems with Köppen-Geiger

Problem

• Climate depends on more than temperature and precipitation.

ロン (得) (ヨ) (ヨ) (ヨ) (ヨ) の(の)

- Can only resolve land.
- Does not adapt to changing climate.

Climate Biome Clustering

└─Problems with Köppen-Geiger

Problem

- Climate depends on more than temperature and precipitation.
- Can only resolve land.
- Does not adapt to changing climate.
- The cut-offs in model are, to some extent, arbitrary.

ロト (得) (日) (日) (日) 日 の(の)

Climate Biome Clustering

└─Problems with Köppen-Geiger

Problem

- Climate depends on more than temperature and precipitation.
- Can only resolve land.
- Does not adapt to changing climate.
- The cut-offs in model are, to some extent, arbitrary.
- No universal agreement to how many classes there should be.

ロン (得) (ヨ) (ヨ) (ヨ) (ヨ) の(の)

Climate Biome Clustering

└─Problems with clustering

Problem

• Dependence on algorithm of choice and hyperparameters.

ロン (得) (ヨ) (ヨ) (ヨ) (ヨ) の(の)

Climate Biome Clustering

Problems with clustering

Problem

• Dependence on algorithm of choice and hyperparameters.

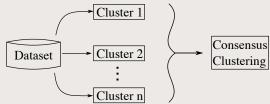


Figure: Many clusterings combined into a single **consensus clustering**.

ロト (得) (日) (日) (日) 日 の(の)

- Climate Biome Clustering
 - Problems with clustering

Problem

• Dependence on algorithm of choice and hyperparameters.

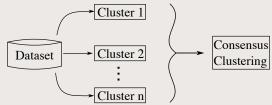


Figure: Many clusterings combined into a single **consensus clustering**.

ロト (得) (日) (日) (日) 日 の(の)

■ Clustering ill-posed - lack measurement of "trust".

- Climate Biome Clustering
 - Problems with clustering

Problem

■ Dependence on algorithm of choice and hyperparameters.

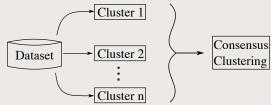


Figure: Many clusterings combined into a single **consensus clustering**.

- Clustering ill-posed lack measurement of "trust".
- Dependence on "hidden parameters" scale of data.

ロト (得) (日) (日) (日) 日 の(の)

- Climate Biome Clustering
 - └─Proposed Solution

Solution

Leverage discrete wavelet transform to classify across a multitude of scales.

- Climate Biome Clustering
 - └─Proposed Solution

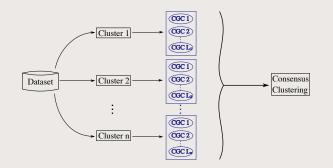
- Leverage discrete wavelet transform to classify across a multitude of scales.
- **2** Use information theory to discover most important scales to classify on.

- Climate Biome Clustering
 - └─Proposed Solution

- Leverage discrete wavelet transform to classify across a multitude of scales.
- **2** Use information theory to discover most important scales to classify on.
- **3** Taking these scales, combine classifications to produce a **fuzzy** clustering that assess the trust at each point.

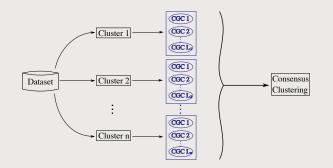
- Climate Biome Clustering
 - └─Proposed Solution

- Leverage discrete wavelet transform to classify across a multitude of scales.
- **2** Use information theory to discover most important scales to classify on.
- **3** Taking these scales, combine classifications to produce a **fuzzy** clustering that assess the trust at each point.



- Coarse-Grain Clustering (CGC)
 - └─Proposed Solution

- Leverage discrete wavelet transform to classify across a multitude of scales.
- **2** Use information theory to discover most important scales to classify on.
- **3** Taking these scales, combine classifications to produce a fuzzy clustering that assess the trust at each point.



<u>Coarse</u>-Grain Clustering (CGC)

└─Results - Effect of Coarse-Graining

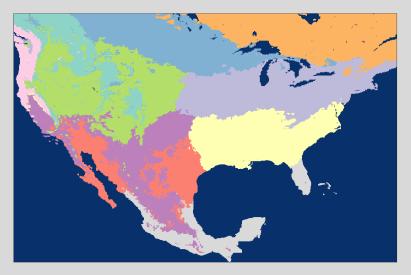


Figure: CGC: K-means $k = 10, (\ell_s, \ell_t) = (1, 1)$

<u>Coarse</u>-Grain Clustering (CGC)

Results - Effect of Coarse-Graining

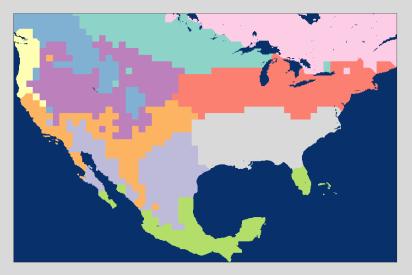


Figure: CGC: K-means $k = 10, (\ell_s, \ell_t) = (4, 1)$

(個人・モン・モン ヨーのへの

<u>Coarse</u>-Grain Clustering (CGC)

└─Results - Effect of Coarse-Graining

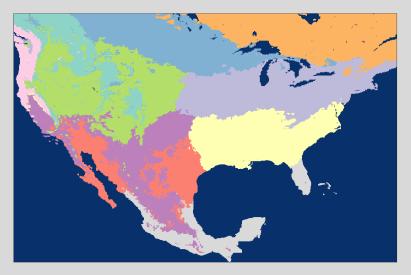


Figure: CGC: K-means $k = 10, (\ell_s, \ell_t) = (1, 1)$

<u>Coarse</u>-Grain Clustering (CGC)

└─Results - Effect of Coarse-Graining

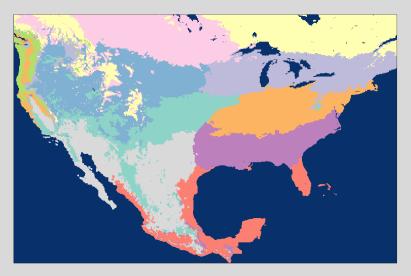


Figure: CGC: K-means $k = 10, (\ell_s, \ell_t) = (1, 6)$

日本・モン・モン ヨー シへで

<u>Coarse</u>-Grain Clustering (CGC)

└─Results - Effect of Coarse-Graining

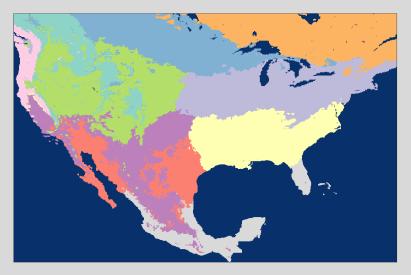


Figure: CGC: K-means $k = 10, (\ell_s, \ell_t) = (1, 1)$

<u>Coarse</u>-Grain Clustering (CGC)

└─Results - Effect of Coarse-Graining

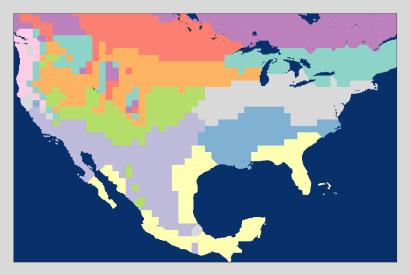
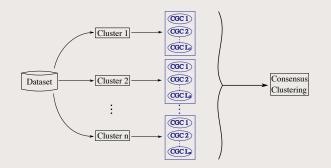


Figure: CGC: K-means $k = 10, (\ell_s, \ell_t) = (4, 6)$

(日本・コン・ヨン ヨー シへで)

- Coarse-Grain Clustering (CGC)
 - └─Proposed Solution

- Leverage discrete wavelet transform to classify across a multitude of scales.
- **2** Use information theory to discover most important scales to classify on.
- **3** Taking these scales, combine classifications to produce a fuzzy clustering that assess the trust at each point.



- <u>Coarse</u>-Grain Clustering (CGC)
 - Lesults Example for K-means K=10



(b)
$$(\ell_s, \ell_t) = (2, 4)$$



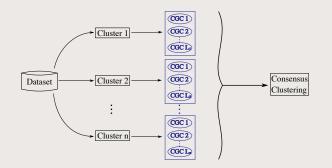
(c)
$$(\ell_s, \ell_t) = (3, 5)$$



(d) $(\ell_s, \ell_t) = (4, 4)$

- Coarse-Grain Clustering (CGC)
 - └─Proposed Solution

- Leverage discrete wavelet transform to classify across a multitude of scales.
- **2** Use information theory to discover most important scales to classify on.
- **3** Taking these scales, combine classifications to produce a fuzzy clustering that assess the trust at each point.



Coarse-Grain Clustering (CGC)

└─Results - Example for K-means K=10

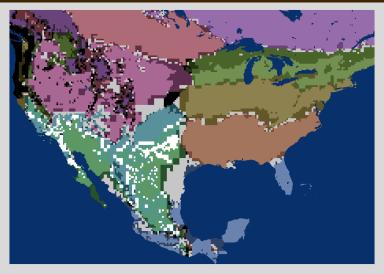
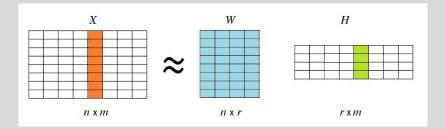


Figure: Consensus clustering from reduced ensemble of clusters for k=10, along with the trust. Grey = multi-class. Darker hue = lower trust.

L Tensor factorizations

∟_{Background}



ロト (日本)(日本)(日本) 日本の文化

L Tensor factorizations

∟_{Background}

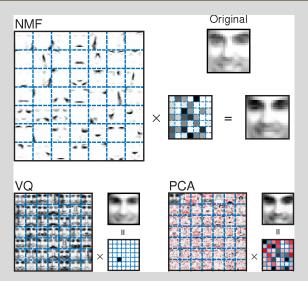
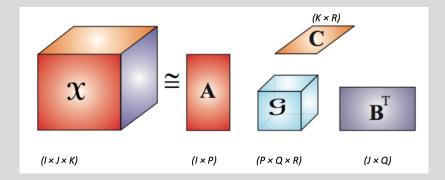


Figure: NMF versus other matrix decompositions (Lee, Seung)

ロ > < 同 > < 三 > < 三 > 三 - シへで

-Tensor factorizations

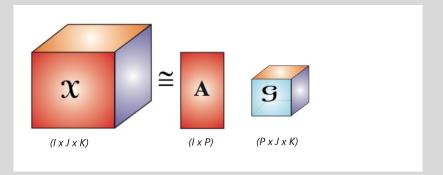
∟_{Background}



ロト (日) (三) (三) (三) () ()

-Tensor factorizations

∟_{Background}



ロト (日) (三) (三) (三) () ()

L Tensor factorizations

L_Background

Problem

• Increasing the number of hidden variables reduces reconstruction error

ロン (得) (ヨ) (ヨ) (ヨ) (ヨ) の(の)

L Tensor factorizations

∟_{Background}

Problem

• Increasing the number of hidden variables reduces reconstruction error

ロン (得) (ヨ) (ヨ) (ヨ) (ヨ) の(の)

More hidden variables is harder to interpret

L Tensor factorizations

∟_{Background}

Problem

- Increasing the number of hidden variables reduces reconstruction error
- More hidden variables is harder to interpret
- At a certain point, one is fitting noise and not signal

ロン (得) (ヨ) (ヨ) (ヨ) (ヨ) の(の)

Summary

• NTF is finding interpretable climate signals



(日本・モン・モン・モー のへの

Summary

- NTF is finding interpretable climate signals
- As seen with clustering, scale is playing a role that we need to analyze



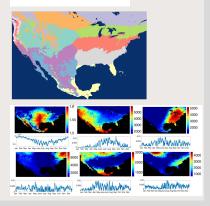


Summary

- NTF is finding interpretable climate signals
- As seen with clustering, scale is playing a role that we need to analyze

• Can we discover latent signatures of El Nino/La Nina?





- Extra Slides
 - -More Tensor Factorizations

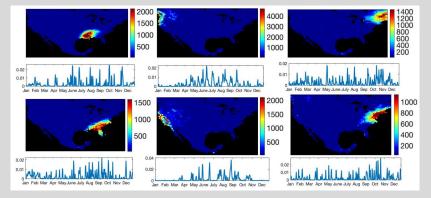


Figure: 1982 Precipitation Modes

(月) (ヨ) (日)

> = ~ ~ ~

- -Extra Slides
 - └ More Tensor Factorizations

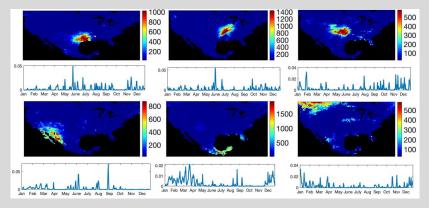


Figure: 1982 Precipitation Modes

- Extra Slides
 - └ More Tensor Factorizations

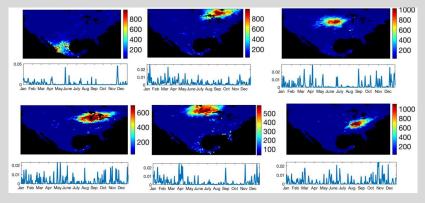


Figure: 1982 Precipitation Modes

ロン (得) (ヨ) (ヨ) (ヨ) ヨー のへの

Extra Slides

└_More Tensor Factorizations

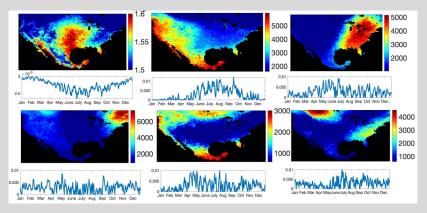


Figure: 1982 Temperature Modes

- Extra Slides
 - └ More Tensor Factorizations

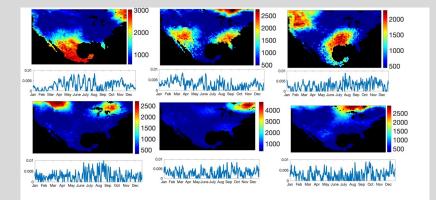
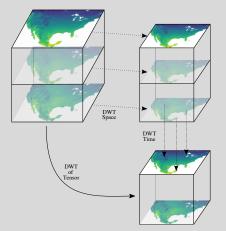


Figure: 1982 Temperature Modes

Extra Slides

└─Discrete Wavelet Transform

- The DWT splits a signal into high and low frequency
- Low temporal signal captures climatology (seasons, years, decades), while low spatial signal captures regional features(city, county, state).



ロト (得) (ヨ) (ヨ) (ヨ) () ()

- Extra Slides
 - -Discrete Wavelet Transform

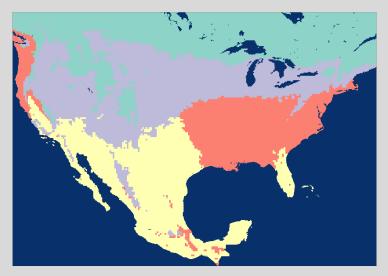


Figure: CGC: K-means k = 4, $(\ell_s, \ell_t) = (2, 3)$

個××注××注× 注 のへで

- Extra Slides
 - Discrete Wavelet Transform

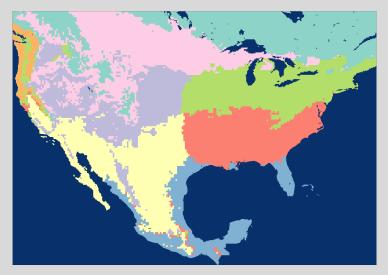


Figure: CGC: K-means k = 8, $(\ell_s, \ell_t) = (2, 3)$

・ イヨ・ イヨ・ ヨー のへで

- _Extra Slides
 - Discrete Wavelet Transform

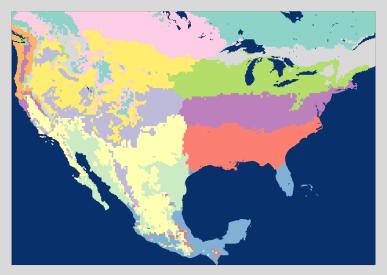


Figure: CGC: K-means $k = 12, (\ell_s, \ell_t) = (2, 3)$

日本・モン・モン ヨー シへで

- _Extra Slides
 - -Discrete Wavelet Transform



Figure: CGC: K-means k = 16, $(\ell_s, \ell_t) = (2, 3)$

NACH REALENCE