When machine learning deciphers the 'language' of atmospheric air masses

Davide Faranda, Lucas Fery, Flavio Pons
Berengère Podvin, Berengère Dubrulle

Lucas.Fery@lsce.ipsl.fr
Davide.faranda@cea.fr
A meteorologist interpret everyday weather by looking at the motion of cyclones (low pressure system) and anticyclones (high pressure systems). They bring stable and dry or wet rainy weather.

To understand climate, we need to read the weather maps for several decades, and categorize cyclones and anticyclones, their position & their frequency. This task is beyond human capabilities, so we use machine learning techniques as an artificial intelligent meteorology.
Tradition statistical **techniques to decompose automatically the weather dynamics** into simple elements work differently than human brain.

They associate to each map, another map that contains a mixture of cyclones and anticyclones.

⇒ **Lack of interpretability**

⇒ **We use the Latent Dirichlet Allocation** that works as a human meteorologist

Van der Wiel et al. 2019
Latent Dirichlet Allocation (LDA)

Generative probabilistic model used in linguistics as a topic model

<table>
<thead>
<tr>
<th>CORPUS</th>
<th>TOPICS</th>
<th>MOTIFS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Document 1</td>
<td>Colors</td>
<td>Mediterranean anticyclone</td>
</tr>
<tr>
<td>Document 2</td>
<td>Fruits</td>
<td>Azores anticyclone</td>
</tr>
<tr>
<td>Document 3</td>
<td>Seasons</td>
<td>Icelandic Low</td>
</tr>
</tbody>
</table>

Blei et al. (2003)
Frihat et al. (2020)
Latent Dirichlet Allocation (LDA)  
Blei et al. (2003)

Generative probabilistic model used in linguistics as a **topic model**

<table>
<thead>
<tr>
<th>CORPUS</th>
<th>Document 1</th>
<th>Document 2</th>
<th>Document 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3 Blue</td>
<td>4 Winter</td>
<td>1 Spring</td>
</tr>
<tr>
<td></td>
<td>1 Red</td>
<td>1 Summer</td>
<td>1 Orange</td>
</tr>
<tr>
<td></td>
<td>2 Orange</td>
<td>1 Autumn</td>
<td>3 Strawberry</td>
</tr>
<tr>
<td></td>
<td>4 Strawberry</td>
<td>2 Orange</td>
<td>2 Apple</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Documents = bags of words
- Words belong to a **finite vocabulary** $V = \{w_i, i \in [1,n]\}$
- Documents represented by vectors $d^l$
- $d^l_i \in \mathbb{N}$ : number of occurrences of $w_i$ in document $l$
Latent Dirichlet Allocation (LDA) by Blei et al. (2003)

Generative probabilistic model used in linguistics as a **topic model**

**CORPUS**

<table>
<thead>
<tr>
<th>Document 1</th>
<th>Document 2</th>
<th>Document 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 Blue</td>
<td>4 Winter</td>
<td>1 Spring</td>
</tr>
<tr>
<td>1 Red</td>
<td>1 Summer</td>
<td>1 Orange</td>
</tr>
<tr>
<td>2 Orange</td>
<td>1 Autumn</td>
<td>3 Strawberry</td>
</tr>
<tr>
<td>4 Strawberry</td>
<td>2 Orange</td>
<td>2 Apple</td>
</tr>
</tbody>
</table>

**TOPICS**

<table>
<thead>
<tr>
<th>Colors</th>
<th>Fruits</th>
<th>Seasons</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4 Blue</td>
<td>0.4 Strawberry</td>
<td>0.4 Winter</td>
</tr>
<tr>
<td>0.3 Red</td>
<td>0.2 Apple</td>
<td>0.2 Summer</td>
</tr>
<tr>
<td>0.2 Orange</td>
<td>0.3 Orange</td>
<td>0.2 Autumn</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2 Spring</td>
</tr>
</tbody>
</table>
Moreover, documents are represented as **mixtures of topics**:

**Document 1**

3 Blue
1 Red
2 Orange
4 Strawberry

**LDA**

**Document-topic distribution**

- 0.7 Colors
- 0.3 Fruits

**Latent Dirichlet Allocation (LDA)**

Blei et al. (2003)
Latent Dirichlet Allocation (LDA) (Blei et al. 2003)

- Document 1
- Document 2
- Document 3
- Document 4
- Document 5
- Document 6
- Document 7

Words occurrences:
- Word 1
- Word 2
- Word 3
- Word 4
- Word 5
- Word 6
- Word 7
- Word 8
- Word 10
Latent Dirichlet Allocation (LDA) 

Blei et al. (2003)

- Document 1
- Document 2
- Document 3
- Document 4
- Document 5
- Document 6
- Document 7

- Topic 1
- Topic 2
- Topic 3

- Word 1
- Word 2
- Word 3
- Word 4
- Word 5
- Word 6
- Word 7
- Word 8
- Word 10

Document-topic distribution

Topic-word distribution

TOPICS
Analogy proposed by Frihat et al. (2020):

<table>
<thead>
<tr>
<th>Linguistics</th>
<th>Physics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Documents</td>
<td>Snapshots (gridded maps)</td>
</tr>
<tr>
<td>Words</td>
<td>Grid-points</td>
</tr>
<tr>
<td>Number of occurrences</td>
<td>Grid-points values</td>
</tr>
<tr>
<td>Topics</td>
<td>Spatial patterns (named <em>motifs</em>)</td>
</tr>
</tbody>
</table>

Grid-point values are converted to positive integers via rescaling, digitization and thresholding.

Application of LDA to gridded observables

Here applied to daily sea level pressure anomaly (NCEP/NCAR reanalysis) over North Atlantic

Need to convert our **observable values** to **positive integers**:

Problem: anomaly can be >0 or <0

Solution: **Split our maps in two maps** (double the ‘vocabulary’ size)
Selection of the number of motifs

The area of the motifs converges to the **typical size of cyclones and anticyclones** \((R \sim 1000 \text{ – } 1500 \text{ km})\)
Selection of the number of motifs

Relative covered area reaches 99% at 28 motifs
We choose to set $N$ as 28

Further motivated by independent analysis on the same dataset:

28 is the upper bound of the number of degrees of freedom obtained computing the local attractor dimensions.

Motifs in daily pressure maps NCEP 1948-2020

- H1
- H2
- H3
- H4
- H5
- H6
- H7
- H8
- H9
- H10
- H11
- H12

- L1
- L2
- L3
- L4
- L5
- L6
- L7
- L8
- L9
- L10
- L11
- L12
- L13
- L14
- L15
- L16

- Azores anticyclone
- Siberian High
- Icelandic Low
- Genoa Low

Weight range: -0.02 to 0.02
Extreme events

We use EM-DAT where European Extreme events are referenced based on their impacts (death toll, victims, state of emergency…) and not on statistical analysis

• Can these events be related to the motifs defined with LDA?
• Are there Motifs that are precursors of extreme events?

Focus on Cold Spells over Europe
Extreme events: cold waves
During the past 70 years, strong increase of anticyclonic patterns over the Mediterranean sea.
Conclusions

• LDA adapted and applied to gridded atmospheric observable (sea level pressure anomaly)

• Identification of relevant localized patterns: cyclones and anticyclones

• Identification of few relevant motifs and precursors for extreme events

• Allow to detect trends in patterns recurrence and intensity
Applications open: Summer school in Trieste

Artificial Intelligence for Detection and Attribution of Climate Extremes

20 June - 2 July 2022
An ICTP Hybrid Meeting
Trieste, Italy

During the last 5-10 years, a large number of extreme weather and climate events in Europe and worldwide have occurred, causing damage to infrastructure and casualties especially in developing countries. This has raised the question about the role of climate change in altering the odds or the magnitude of a number of such events and the new “science of attribution” has began with several attribution published all around the world. The aim of the school is to define techniques to tackle the problem of attributing meteorological extreme events to climate change by means of machine learning technologies. Lectures will also focus on determining causal links of extreme events with the underlying climate dynamics as the atmospheric circulation. The school will also discuss and provide the bases for communicating attribution results to the general public, stakeholders and other scientists in an exact although non specialist language.

Directors:
1. BEVACqua, Helmholtz Centre for Environmental Research – UFZ, Germany
2. COPPOLA, ICTP, Italy
3. COUNIOOL, Vrije Universiteit Amsterdam, Netherlands
4. FARAONDE, LSCE-IPSL, CNRS, France
5. JERICHO, ENS Paris, France
6. VAITARD, LSCE-IPSL, CNRS, France
7. VRAC, LSCE-IPSL, CNRS, France
8. YIOU, LSCE-IPSL, CEA, CNRS, France

Local Organiser:
1. COPPOLA, ICTP, Italy

Speakers:
1. BARNES, Colorado State U, USA
2. BEVACqua, UZ Leipzig, Germany
3. CAMPS-VALLS, IEP-UVED, Spain
4. COPPOLA, IPCC, Italy

Further information:
http://info.ictp.it/event/98002/
sm371@ictp.it
Thank for your Attention

To keep in touch

mail: davide.faranda@lsce.ipsl.fr

@DaviFaranda

Any Questions?
References


• EM-DAT database [https://www.emdat.be/](https://www.emdat.be/)

• Preprint at [https://www.researchsquare.com/article/rs-608588/v1](https://www.researchsquare.com/article/rs-608588/v1)