eScience: more than just digital technology

Romulo Goncalves
Data...
Data...
Data...
Data...
Data...
Data...
Problem size...
Problem size…
eScience Technology

- Fast data ingestion
- Dynamic step-wise search
- Flexible hypotheses validation
eScience Technology

- Fast data ingestion
- Dynamic step-wise search
- Flexible hypotheses validation
eScience Technology

- Fast data ingestion
- Dynamic step-wise search
- Flexible hypotheses validation
Phenology...

Google Earth Engine
Phenology...

Google Earth Engine
Enabling Science...
Enabling Science...
Enabling Science...
Enabling Science...

```
In [6]: samples = data.reshape((4, -1)).T
out[6]: (13251842, 4)

Kmeans
In [9]: clf = sklearn.cluster.KMeans(n_clusters=12)
labels = clf.fit_predict(samples)
labels.shape
Out[9]: (13251843,)

Plot Kmeans clusters
In [10]: import matplotlib.pyplot as plt
plt.figure(figsize=(20, 10))
plt.imshow(data.reshape((40889, 4587)), cmap="Set3")
Out[10]: <matplotlib.image.AxesImage at 0x7f0e466429e8>
```
Identify regions with similar phenology
Enhancing Science…
Enhancing Science...
Enhancing Science…
Find patterns in spatio-temporal fields

- Find patterns which occur frequently and simultaneously
- SVD on the product of standardized datasets
  - Cross variance matrix

\[ \begin{bmatrix} m \\ n \end{bmatrix} \begin{bmatrix} m \\ n \end{bmatrix} = \begin{bmatrix} m \end{bmatrix} \]
Find patterns in spatio-temporal fields

- Find patterns which occur frequently and simultaneously

- SVD on the product of standardized datasets
  - Cross variance matrix

\[ \mathbf{X}^{m \times n} = \mathbf{m}^{m \times m} \]
Find patterns in spatio-temporal fields

• Find patterns which occur frequently and simultaneously

• SVD on the product of standardized datasets
  – Cross variance matrix

\[
X = \begin{pmatrix}
m \\
n \\
m \\
\end{pmatrix}
\begin{pmatrix}
m \\
m \\
\end{pmatrix}
= \begin{pmatrix}
m \times n \\
m \\
\end{pmatrix}
\begin{pmatrix}
m \\
m \\
\end{pmatrix}
\]
Find patterns in spatio-temporal fields

- Find patterns which occur frequently and simultaneously
- SVD on the product of standardized datasets
  - Cross variance matrix

\[
A^{T}B = [R_{A} \quad R_{B}^{T}] Q_{B}^{T} = Q_{A} C \quad Q_{B}^{T} = Q_{A} \quad \tilde{U}_{C} S_{C} \quad \tilde{V}_{C}^{T} Q_{B}^{T} = U_{C} S_{C} \quad V_{C}^{T}
\]
Find patterns in spatio-temporal fields

- Find patterns which occur frequently and simultaneously

- SVD on the product of standardized datasets
  - Cross variance matrix

\[
X = \begin{bmatrix}
m \\
n \\
m \\
m \\
\end{bmatrix} = \begin{bmatrix}
m \\
m \\
\end{bmatrix}
\]
Find patterns in spatio-temporal fields

- Find patterns which occur frequently and simultaneously
- SVD on the product of standardized datasets
  - Cross variance matrix

\[
\begin{align*}
X &= mn \\
A^{T} &= Q_{A} \begin{bmatrix} R_{A} & R_{B}^{T} \end{bmatrix} Q_{B}^{T} \\
C &= Q_{A} \begin{bmatrix} \tilde{U}_{C} & S_{C} & \tilde{V}_{C}^{T} \end{bmatrix} Q_{B}^{T} \\
U_{C} &= S_{C} V_{C}^{T}
\end{align*}
\]
Not only Scientific results…

• Guidelines
  – For software development

• Findable, Accessible, Interoperable, and Reusable (FAIR)
  – Software
  – Data

• Educational material
  – Tutorials for master courses

• Collaborations
Take home…

• Enable Science
  – Data deluge
  – Problem size increase
  – Digital revolution

• Enhancing Science

• Not only Scientific results

• eScience: more than just digital technology
Questions?

Image from: http://aboutinterviews.com/5-questions-to-not-ask-in-a-job-interview/