P48-Identification and clustering of microbial growth

Project owner: SynBio group
Institute of Technology, University of Tartu

TEAM:
Ülle Püttsepp
Klāvs Jermakovs
Alejandra Duque-Torres

14/12/2020
What?
CELL FACTORIES:
- Microbial cells with **custom genome**
- Producing high-value chemicals

Problem?
Researchers need to characterize organism.
- By **manually studying** Bioreactor data
- **Deriving phases** of microbial growth
- Infeasible for many parallel experiments
The project

Data

- Batch data describing **Biomass** change over **time** (170h)
- **Indirect Biomass** measurements (ODa, every hour)
- Contains **direct** measurements (Biomass, every 20h)

Tool

- **Automates** data pre-processing
- **Clusters** and identifies phases of Growth
How it Works?

Pre-processing

- **ODa into Biomass** through linear regression fit on direct measurements
- Calculates $\mu$ through taking derivative

![Graph showing ODa over time](image-url)
How it Works?

Pre-processing

- **ODa into Biomass** through linear regression fit on direct measurements
- Calculates $\mu$ through taking derivative
How it Works?

Pre-processing

- **ODa into Biomass** through linear regression fit on direct measurements
- Calculates $\mu$ through taking derivative
How it Works?

Pre-processing

- **ODa into Biomass** through linear regression fit on
- Calculates $\mu$ through taking derivative
How it Works?

Clustering

- Transforms **Biomass into log** scale, finds exponential growth regions through KMeans.
- Finds **region $\mu_{\text{max}}$** through iterative $R^2$ optimization.
- Unsupervised **KMeans** algorithm on (Time, Biomass, $\mu$) to cluster + **Silhouette** analysis and **Elbow** gives number of optimal clusters
How it Works?

Clustering

- Transforms Biomass into log scale, finds exponential growth regions through KMeans.
- Finds region $\mu_{max}$ through iterative $R^2$ optimization.
- Unsupervised KMeans algorithm on (Time, Biomass, $\mu$) to cluster + Silhouette analysis and Elbow gives number of optimal clusters.
How it Works?

Clustering

- Transforms **Biomass** into log scale, finds exponential growth regions through KMeans.
- Finds **region $\mu_{max}$** through iterative R2 optimization.
- Unsupervised **KMeans** algorithm on (Time, Biomass, $\mu$) to cluster + **Silhouette** analysis and **Elbow** gives number of optimal clusters
Results

**Precise µ_max**

- Equation: $y = 0.04965 \times x + 0.44646$
- R-squared = 0.99748
- $\mu_{Max} = 0.04965$

**Growth Phases**

- **Phase 1**
- **Phase 2**
- **Phase 3**
- **Phase 4**
Thanks
Gracias
Aitäh!

Questions?

https://github.com/aduquet/Clustering_Microbial-growth

Ülle Püttsepp
Aleja Duque-Torres
Klāvs Jermakovs