Emmanuel Faure[1]
Benoit Lombardot[1]
Miguel Luengo Oroz[2]
Matteo Campana[3]
Nadine Peyriéras[4]
René Doursat[1]
Paul Bourgine[1]

(1) Centre de Recherche en Epistémologie Appliquée, CNRS - Ecole Polytechnique, *{emmanuel.faure,benoit.lombardot,paul.bourgine,rene.doursat}@shs.polytechnique.fr*
(2) BIT - ETSIT , Universidad Politècnica de Madrid *maluengo@die.upm.es*
(3) Informatica e Sistemistica, Università de Bologna. *mcampana@deis.unibo.it*
(4) DEPSN, CNRS - Gif-sur-Yvette *peyrieras@inaf.cnrs-gif.fr*

Active Machine Learning for Embryogenesis

Description

The intrinsic complexity of biological systems creates huge amounts of **unlabeled experimental data.** The exploitation of such data can be achieved by performing active machine learning accompanied by a high-level symbolic expert who defines categories and their best boundaries using as little data as possible. We present a global strategy for designing active machine learning methods suited for the observation and analysis of complex systems, such as embryonic development.
We developed a procedure that uses all available knowledge, whether gathered manually or automatically, and is able to readjust when new data is provided. We show that it is a powerful method for the investigation of the morphogenetic features of embryogenesis. It will make possible to properly reconstruct the in vivo cell morphodynamics, a main challenge of the post-genomic era.

Project

Embryomics is devoted to the morphodynamical "reconstruction" of the **cell lineage tree** underlying the processes of animal embryogenesis.

Our goal is to fully **reconstruct the dynamics of cell divisions and movements** from time-lapse series of high-resolution optical sections obtained by multiphoton laser scanning microscopy throughout embryonic development of live animals .

We design a set of strategies, methods and algorithms to "sequence" the cell lineage tree as a **branching process** annotated in space and time.

Embryomics will allow the **automated tracking** of events such as cell division and cell death in live embryos and give us access to parameters such as the rate of cell proliferation in time and space.

This kind of data is highly relevant to investigate stem cell populations, early steps of cancerogenesis and drug effects in vivo.
A comparative Embryomics strategy is the best way to evaluate the relevance of animal models, whatever their evolutionary distance to human, for further medical applications.

DATA Treatment

4D Embryos

~ 10 Embryos
~ 24 h / Embryo
~ Δt =1mn
~ 2500 cells/ time steps
⇒Number of cells ≅ $5x10^8$

→ Filtering → Detection → Segmentation → Tracking →

Parameters

Nuclei & Membrane :
→Shape, Volume
→Velocity
→Neighbors
→Intensity
→...

Given $\{x_i\}$ a cells list where each $x \in X \subset \mathbb{R}^d$ is a d-dimensional vector, we denoted by $x_i = \{p_1, ..., p_d\}$ the characteristics list of d parameters $p \in$ P .The complete characteristics space P that we used is of dimension $d = 42$.

Unlabeled Data

Cognitive Perception — Active Learning — Embryomics

Data : Plants,trees,flowers, etc..
Dimension : size, color, trunk, foliage, lifespan,...
Goal : find the categorises of each species

Classification Function

Estimation a classification function $f : \mathbb{R}^d \rightarrow \{1, \ldots, N_C\}$ where N_C is the number of classes

Two interrelated classifiers :
- (f_S) for segmentation classifier with { - $f_S(x_i) = -1$ for an incorrect segmentation ; - $f_S(x_i) = +1$ for a correct segmentation .

- (f_M) classified the correctly segmented cells, and only these, into 7 classes corresponding to the different phases of cell mitosis.

Perceptive Categorization

Unsupervised Learning

Density Clustering Method done by Support vector machines (SVM) specifically One-Class SVM clustering , to put all the data of one class into a hypersphere.

A first (f_S) classification was performed with a density clustering method.

Concept Acquisition

Common knowledge :
Tree :
- size limits
- big trunk
- ..

Flowers :
- many colors
-...

Supervised Learning

Human supervisor using prior knowledge

The supervisor specified a set of constraints $K = \{k_1, ..., k_N\}$ one for each class, calculated on parameters **P.**
Constraint k_i represented the boundaries of a specific partition in **P** space.
A element x_i was thus put in class j (i.e., $f(x) = j$) if it satisfied constraint k .

Readjust the previous (f_S) classification with constrainsts :
- epithelial cells are elongated
- min and max cell volume limits
- Intensity Limit for Detection

Initialized (f_M)
- **Prophase** : membrane deformation from elliptic to spheric
- **Prometaphase** : changement in nucleus volume and convexity
- **Metaphase** : Planar Arrangement of the chromosomes in the future cell division plane
- **Anaphase** : Separation of the two sets of chromosomes yielding two nuclei within a single cell membrane
- **Telophase** : Membrane constriction in the cell division plane
- **Cytokinesis** : Separation of daughter cells

Intervention

Ask for ambiguous elements

Semi-supervised Learning

1 - Calculate elements at the boundaries (support vectors) by a Multi-Class SVM classifier.

2 - Manual annotation of elements at the boundaries
-> according to previous automatic classification
→ Nothing happens
-> reevaluation of a support vector that was incorrectly classified
→ Recalculate the new elements at the boundaries

=> moving the boundaries toward their optimal position.
=> Increase the quality of the reconstruction

Repeat all these steps until the supervisor agrees with all these classification.
Convergence to the best margin between classes.

Tools for Annotation

Expertise

Ask if any doubt

Self-supervised Learning

Add the new elements in the curent classification and automatically classify them.

Recalculate the new elements at the boundaries and use Semi-supervised learning to annotate new elements at the boundaries

Can learn alone new elements

A new 3D image (New time steps or new embryos) was added and categorized with the knowledge of the previous Multi-Class SVM classification.

Classification could be readjusted by Manual Annotation

Thanks to Jonathan Platkiewicz and Carine Mantoulan

Bibliography

S.G. Megason ans S.E. Fraser, Digitizing life at the level of the cel l : high-performance laser-scanning microscopy and image analysis for in toto imaging of development,Mech.Dev. (2003), 1407-1420.
J. F. Bowring, J. M. Rehg, and M. J. Harrold. Active learning for automatic classification of software behavior. International Symposium on Software Testing and Analysis, pages 195205, 2004.
V.N. Vapnik, The nature of statiscal learning theory, Springer-Verlag New York, Inc., New York, NY, USA, 1995.
B. Scholkopf, J.C. Platt, J. Shawe-Taylor, A.J. Smola, and R. C. Williamson. Estimating the support of a high-dimensional distribution. Neural Computation, 2001.
K. Bennett and A. Demiriz. Semi-supervised support vector machines. In Ad-vances in Neural Information Processing Systems 11, 1998.

Embryomics NEST Adventure 012916