Caenorhabditis elegans lifespan prediction from early adulthood health data with a hidden Markov model

Sangbin Han[1]

[1]School of Biological Sciences, Seoul National University, Seoul 151-747, Korea

Background

- ***Caenorhabditis elegans* (*C. elegans*)** is a transparent nematode (~ 1 mm in length) which has been widely used as a model organism in biology.
- It has a short life cycle of about 3 days and an average lifespan of 2-3 weeks.
- It is the first model organism for which we have a complete cell lineage, a complete connectome (map of neuronal connections), and a complete genome sequence.
- Due to the above advantages, *C. elegans* has been a prominent model organism for studying aging.

Science magazine cover image (24 December 2010)

Research Goal

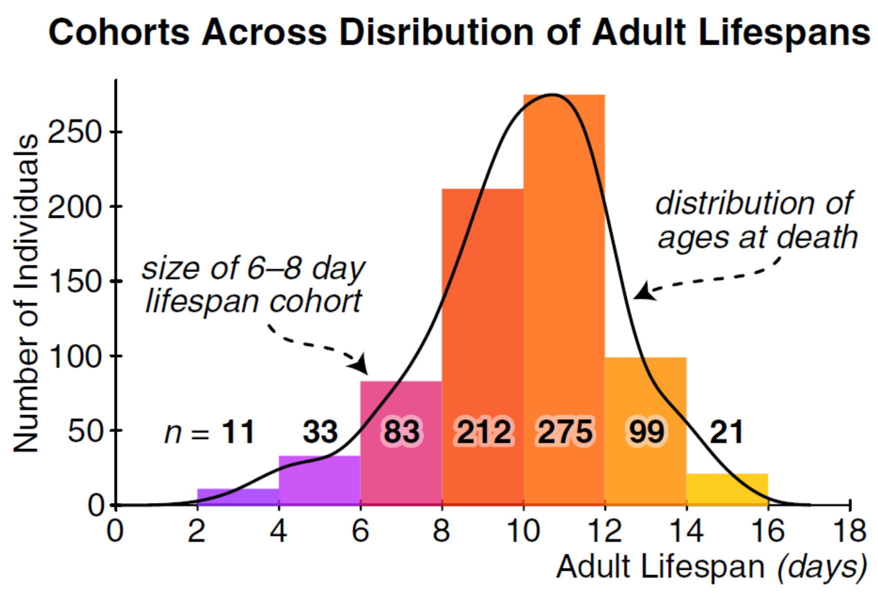
- To construct a computational model that predicts *C. elegans* lifespan from early adulthood health data sequence.
  - ✓ Early selection of long-lived or short-lived *C. elegans* can assist in the longitudinal analysis of aging.

Experimental Design

Data Acquisition and Processing

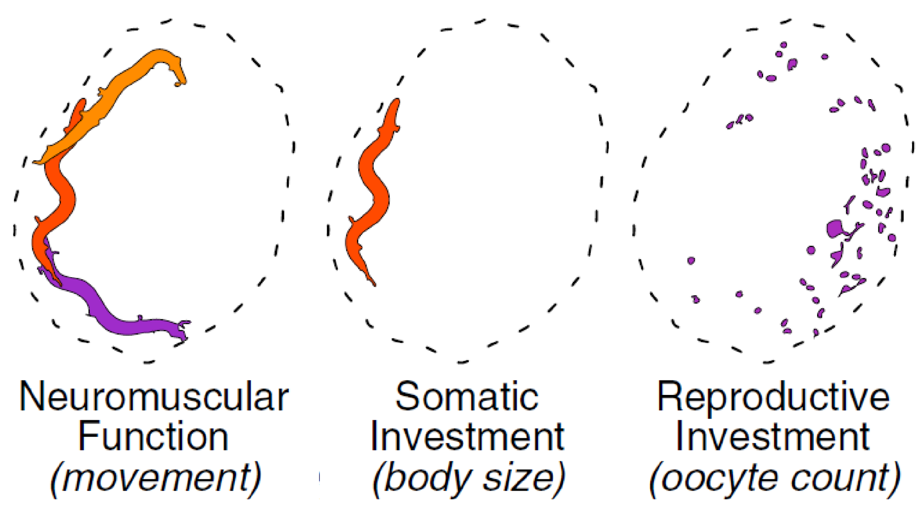
- Zhang et al., 2016, *Cell Systems*
- 734 isogenic individuals kept in identical environment
- Longitudinal physiological measures by lifelong high-resolution imaging (every 3 hours over their lifespans)

Early adulthood (0-2 days) measures (17 per individual) were used.



(Zhang et al., 2016, *Cell Systems*)

1. Neuromuscular function
   - Displacement over 3 hours (mm)
   - Little (33.3%): < 0.438 mm
   - Normal (33.2%): < 0.545 mm
   - Large (33.5%): ≥ 0.545 cm
2. Somatic investment
   - Cross-sectional size (mm^2)
   - Small (33.3%): < 0.0666 mm^2
   - Normal (33.5%): < 0.0809 mm^2
   - Large (33.2%): ≥ 0.0809 mm^2
3. Reproductive investment
   - Cumulative area of eggs laid (mm^2)
   - Small (33.4%): < 0.0370 mm^2
   - Normal (33.3%): < 0.0775 mm^2
   - Large (33.3%): ≥ 0.0775 mm^2



(Zhang et al., 2016, *Cell Systems*)

- Individuals were classified into 3 groups according to their adult lifespans (days).
  1. Short-lived individuals (n = 243): < 9 days 9 hours
  2. Normal-lived individuals (n = 234): < 11 days 3 hours
  3. Long-lived individuals (n = 257): ≥ 11 days 3 hours

Hidden Markov Model

Number of models = 3 (short, normal, or long-lived)

- Number of hidden states (n) = [1, 4, 7, 10, 13, 16, 19]
- Number of observable states (early adulthood health measures) = 27 states = 3 (movement) × 3 (body size) × 3 (area of eggs laid)

$$\frac{1}{n+1} I_n + \begin{bmatrix} \frac{1}{n+1} & \cdots & \frac{1}{n+1} \\ \vdots & \ddots & \vdots \\ \frac{1}{n+1} & \cdots & \frac{1}{n+1} \end{bmatrix}$$

Initial transition matrix

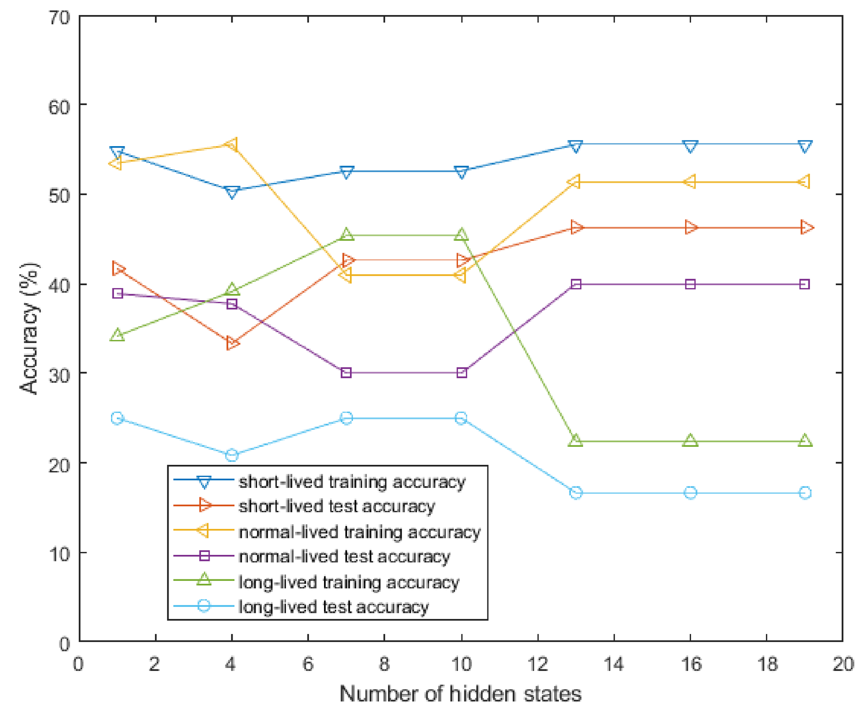
$$\begin{bmatrix} \frac{1}{n} & \cdots & \frac{1}{n} \\ \vdots & \ddots & \vdots \\ \frac{1}{n} & \cdots & \frac{1}{n} \end{bmatrix}$$

Initial emission matrix

Datasets were randomly splitted into training (50%) and test (50%) sets.
Example sequence of emissions:
[2, 5, 15, 15, 15, 15, 15, 27, 27, 27, 27, 27, 27, 18, 18, 18, 18]

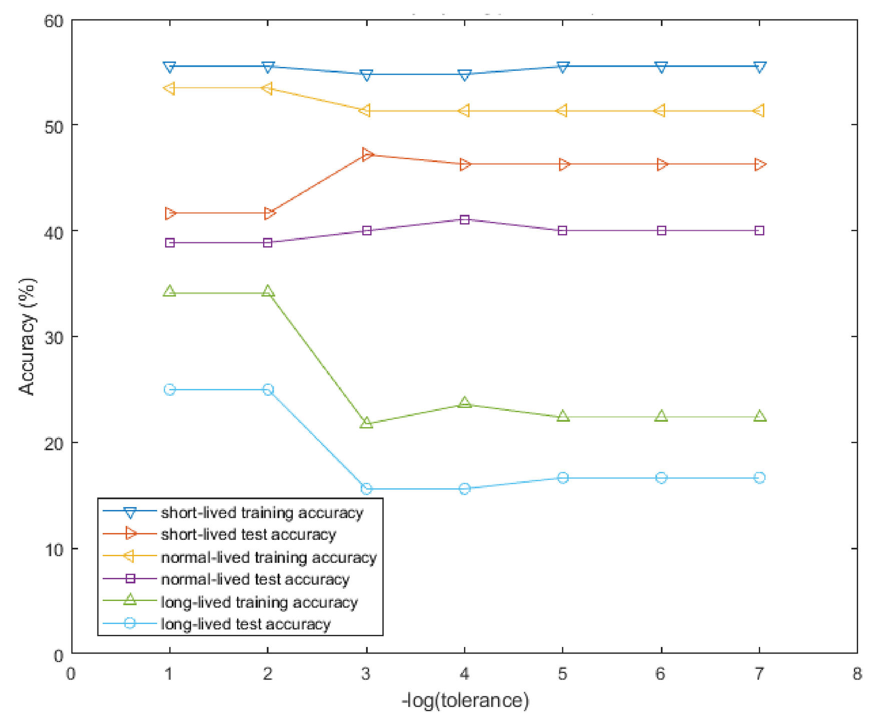
Results

Accuracy by number of hidden states



- Tolerance = 1e-6
- Max iterations = 100
- 13 hidden states showed better test accuracy.
- Short-lived training: 55.56%
- Short-lived test: 46.30%
- Normal-lived training: 51.39%
- Normal-lived test: 40%
- Long-lived training: 22.36%
- Long-lived test: 16.67%

Accuracy by –log(tolerance)



- Number of hidden states = 13
- Max iterations = 100
- 1e-4 tolerance showed better test accuracy.
- Short-lived training: 54.81%
- Short-lived test: 46.30%
- Normal-lived training: 51.39%
- Normal-lived test: 41.11%
- Long-lived training: 23.60%
- Long-lived test: 15.63%

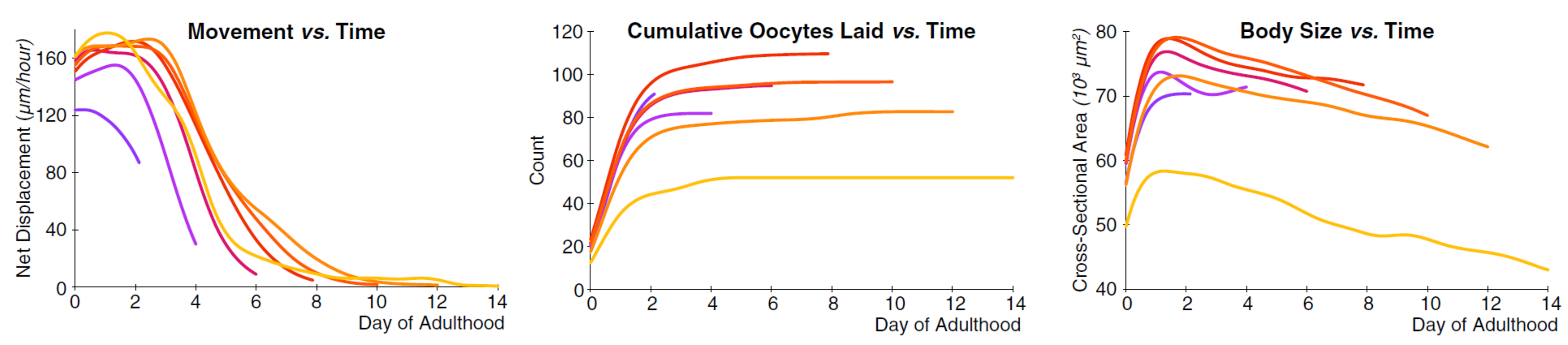
Discussion

Problems

- Low prediction accuracy for lifespans of long-lived individuals
- Below half test set prediction accuracy

Possible Reasons

1. Initialization of transition matrix and emission matrix
   - Transition matrix and emission matrix were arbitrarily initialized.
   - We might make a much better model with various initializations of the two matrices.
2. Same model architecture for all model
   - Same number of hidden states, same tolereance, and same maximum number of iterations were used.
3. Low number of samples
4. Early adulthood health in *C. elegans* is not related to aging (maybe not)



(Zhang et al., 2016, *Cell Systems*)

5. Inappropriate classification of *C.elegans* individuals
   - There might be a subclass of *C.elegans* which has a different longevity mechanism.
   - Clustering techniques (e.g. principal component analysis) can be applied.

References

- Kenyon, Cynthia. "The nematode Caenorhabditis elegans." *Science* 240.4858 (1988): 1448-1453.
- Brenner, Sydney. "The genetics of Caenorhabditis elegans." *Genetics* 77.1 (1974): 71-94.
- Izquierdo, Eduardo J., and Randall D. Beer. "The whole worm: brain–body–environment models of C. elegans." *Current opinion in neurobiology* 40 (2016): 23-30.
- Zhang, William B., et al. "Extended twilight among isogenic C. elegans causes a disproportionate scaling between lifespan and health." *Cell systems* 3.4 (2016): 333-345.