

# Prediction of Growth Parameters of In Vitro Microtubules in Association with Different Isoforms of Tau

Arnab Bhattacharya  
arnab@cs.ucsb.edu

Ambuj K. Singh  
ambuj@cs.ucsb.edu

Department of Computer Science, University of California, Santa Barbara, CA 93106, USA.

## ABSTRACT

*Microtubules are cellular organelles that are important for many cellular processes including cell shape and cell division. Microtubules grow and shorten by a process called the dynamic instability. The dynamics of microtubules can be regulated by different microtubule associated proteins such as tau. Predicting the behavior of microtubules for a hitherto unseen condition of tau from a database of already performed experiments with other conditions of tau will guide the biologists in performing future experiments and exploring unusual values.*

*We employ two different methods of prediction – nearest neighbor and surface fitting. The nearest neighbor method first looks at the neighbors of a condition and its relationship with them. It then applies the average change of all similar relationships in the database to the neighbors to predict. The surface fitting method finds the equation of a surface that best describes the database values and then uses that surface to predict the value for the unseen condition. Experiments were performed with in vitro data with different variants of tau. Since growth is the most important phenomenon for this type of data, two growth parameters were predicted: i) the median growth rate of the microtubules and ii) the percentage of time the microtubules grow. The relative differences between the predicted and the actual values were at most 2%.*

## 1. MICROTUBULES

Microtubules are long, hollow, cylindrical structures found within cells, about 25nm in diameter and up to several microns long. They are polymers of tubulin and are formed by assembly of  $\alpha$  and  $\beta$  subunits. The microtubules are part of the cytoskeleton, but in addition to structural support they are useful for many other processes, including cell division, where they attach to the chromosomes in order to segregate them correctly [2]. Microtubules grow and shorten by a process called the *dynamic instability*.

## 2. DYNAMICITY OF MICROTUBULES AND TAU

The dynamic behavior of microtubules changes when the cell is injected with microtubule associated proteins such as tau and its various isoforms. This change in dynamicity is extremely important biologically. For example, the dysfunction of tau has been correlated with a variety of neuro-degenerative diseases, including Alzheimer disease, fronto-temporal dementia with Parkinsonism associated with chromosome 17 (FTDP-17), Pick disease, and progressive supranuclear palsy [3].

To ascertain the growth and shortening behaviors under different conditions, biologists record videos of the microtubules. Microtubules are tracked either manually or automatically and dif-

ferent parameters are measured [3]. Microtubule behavior can be observed under two different conditions – in vivo (living cells are injected with drugs and microtubules in those cells are observed) and in vitro (tau and tubulin concentrations are prepared outside living cells and microtubule movements are observed). The in vivo microtubules almost always attain steady state whereas the in vitro microtubules may be observed in a pre steady state where most of them are growing for most part of the tracking time frame. This study mainly concerns with the in vitro growth of microtubules in association with different concentrations of the protein tau and its different variants.

## 2.1 Tau

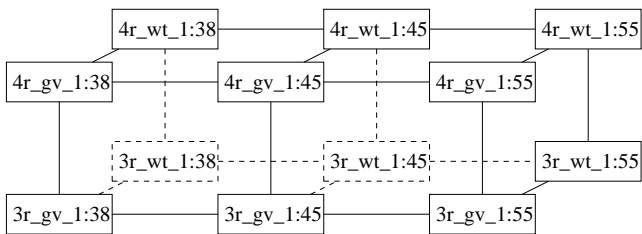
There are two main isoforms of tau—4-repeat tau and 3-repeat tau—formed by alternative splicing from a single tau gene [1]. Each of the isoforms possess either 4 or 3 imperfect 18 amino acid repeats separated by 13-14 amino acid inter-repeats. 3R tau lack the first inter-repeat and the second repeat. The variant of tau that occurs most frequently in nature is called the *wild-type* tau. Mutants such as the G272V (the 272nd amino acid is changed from glycine to valine) are also found naturally. Other than the type and the variant of tau, another important property for in vitro studies is the relative concentration of tau with respect to tubulin. Thus, each tau condition has 3 attributes – 1) the type of the tau protein, 2) the variant of tau (wild-type or mutant) and 3) the concentration of tau with respect to tubulin.

## 3. PREDICTION

Gathering important observations for a particular condition of tau requires considerable time and effort on the part of the biologists. Tau and tubulin have to be prepared, microtubule videos have to be recorded and the microtubules have to be tracked (often manually) before any analysis can be done. Predicting the outcome of an experiment without actually performing it, is therefore, very useful. Prediction will guide the biologists in planning their future experiments from a large choice of possible candidates. It may be interesting to explore a particular condition for which an unusual value has been predicted before experimenting with other conditions. In this section, we describe two methods, nearest neighbor and surface fitting, that use a database of experiments already performed to predict the outcome of a new experiment.

### 3.1 Nearest Neighbor

Suppose there are  $n$  conditions in the database and each condition has  $m$  attributes. We represent condition  $i$  by  $c_i$ , the value of the parameter of interest of condition  $c_i$  by  $v_i$  and the attribute  $k$  of condition  $i$  by  $c_{ik}$ . We say two conditions  $c_i$  and  $c_j$  are neighbors if and only if  $\exists k$  such that  $c_{ik} \neq c_{jk}$  and  $c_{ik'} = c_{jk'}, \forall k' \neq k$ . The



**Figure 1: The grid of conditions that are used as the database for prediction.  $4r$  and  $3r$  denote 4-repeat tau and 3-repeat tau respectively.  $wt$  and  $gv$  denote wild-type and G272V mutant tau respectively.  $1:38$ ,  $1:45$  and  $1:55$  are different concentrations of tau with respect to tubulin.**

edge relationship from condition  $c_i$  to its neighbor  $c_j$  is denoted by  $e(c_i, c_j) = c_{ik} \rightarrow c_{jk}$  and the edge change is calculated as  $\delta(c_i, c_j) = v_i - v_j$ . To predict the value  $v_x$  for an unknown condition  $c_x$ , first its set of neighbors  $\{c_z\}$  in the database are found. The edge relationship of  $c_x$  to each  $c_z$  is then searched in the database and the average of all such changes from the database are applied to  $v_z$  to predict  $v_x$ . The predicted  $v_x$  from all the neighbors  $c_z$  are averaged once more to determine the final value of  $v_x$ . Thus,

$$v_x = \text{avg}_{c_z} \left\{ v_z + \text{avg}_{(c_i, c_j)} \left\{ \delta(c_x, c_z) | e(c_i, c_j) = e(c_x, c_z) \right\} \right\}$$

Figure 1 shows a database of experimental conditions. Assume that we have not performed any experiment with  $4r\_gv\_1:38$  and we want to predict its value, given the rest of the database.  $4r\_wt\_1:38$  is a neighbor and the relationship with the neighbor is  $gv \rightarrow wt$ . Five similar relationships can be found in the database ( $4r\_gv\_1:45$  to  $4r\_wt\_1:45$ ,  $3r\_gv\_1:38$  to  $3r\_wt\_1:38$ , etc.) Adding the average edge change of all such relationships to the value of  $4r\_wt\_1:38$  will provide one prediction for  $4r\_gv\_1:38$ . The two other neighbors  $3r\_gv\_1:38$  and  $4r\_gv\_1:45$  will provide two more such predictions. The average of all of these predictions is the final predicted value.

This is the most simple way of prediction. Instead of looking at changes for a single attribute, neighbors which differ by two or more attributes can be considered. Also, not all edge relationships may be good for prediction. If for an edge, the differences in the values for two neighbors vary widely, this implies that the edge change does not change the parameter value consistently and is therefore not useful for prediction. If, on the other hand, the changes across an edge are consistent, then this means that the edge affects all the conditions in the same manner and is therefore more likely to predict accurately the value for an unseen condition.

### 3.2 Surface Fitting

The second method or the surface fitting method is global unlike the first one as it takes into account all the values in the database for prediction. This method finds the equation of a surface that best describes the values for the different conditions in the database. First, all the possible values for each attribute are numbered. For example, the different values of tau to tubulin ratio can be numbered as follows:  $1:38$  as 1,  $1:45$  as 2 and so on. Denoting the value of attribute  $k$  of condition  $c_i$  by  $\phi_{ik}$ , we can represent the value  $v_i$  for condition  $c_i$  as a linear combination of its attribute values:

$$v_i = \alpha_0 + \sum_{k=1}^m \alpha_k \phi_{ik} \quad (1)$$

A least squares regression is performed with all the database conditions in order to learn the coefficients  $\alpha_k$  for the above surface

	Nearest neighbor	Surface fitting
Growth rate	0.064	0.050
Growth time	0.010	0.043

**Table 1: Average errors of predicting the two growth parameters from the two prediction methods described in Section 3. The database of conditions are shown in Figure 1.**

equation. For any unseen condition  $c_x$ , Eq. (1) can then be employed to predict the value of  $v_x$ .

## 4. EXPERIMENTS

Since for in vitro studies, growth is the most important phenomenon of a microtubule, two growth parameters were predicted. The first parameter is the median growth rate of the microtubules and the second one is the percentage of time the microtubules grow.

We employed the two prediction methods described earlier to the database of conditions shown in Figure 1 in a leave-one-out cross-validation manner. One condition was left out of the database and the rest 11 were used to predict its value. The error was measured as the absolute difference of the predicted value to the actual value. This was repeated for all the conditions in turn and the average error was calculated. Note that the possible values for both the parameters were normalized between 0 and 1. Table 1 shows the results. It is interesting to note that the nearest neighbor method which locally looks at the neighbors performs better for time spent in growth while the surface fitting method which takes into account the global view of the database performs better for the growth rate.

## 5. CONCLUSION

The change in dynamic instability of microtubules due to the binding of different variants of tau is important biologically due to implications on various diseases including Alzheimer's. Since experimentation with different isoforms of tau in different concentrations requires a lot of time and effort, predicting the important parameters of microtubule behavior for a new experiment will guide the biologists in performing experiments and exploring unusual values. We designed two methods of prediction, nearest neighbor and surface fitting, and used them to predict two growth parameters, median growth rate and time spent in growth. The average relative error in prediction was less than 2%.

In future, we would like to investigate more into how a local method like the nearest neighbor compares with a global method like the surface fitting. The effects of tau for in vivo experiments and how the parameters change from in vivo to in vitro are worth exploring as well.

**Acknowledgment.** A. C. LeBoeuf, S. F. Levy, Prof. S. C. Feinstein, Prof. L. Wilson, MCDB, UCSB; NSF ITR grant #0331697.

## 6. REFERENCES

- [1] M. Goedert, M. G. Spillantini, R. Jakes, D. Rutherford, and R. A. Crowther. Multiple Isoforms of Human Microtubule-Associated Protein Tau: Sequences and Localization in Neurofibrillary Tangles of Alzheimer's Disease. *Neuron*, 3(4):519–526, 1989.
- [2] J. S. Hyams and C. W. Lloyd. *Microtubules*. Wiley-Liss, 1994.
- [3] S. F. Levy, A. C. LeBoeuf, M. R. Massie, M. A. Jordan, L. Wilson, and S. C. Feinstein. Three- and Four-Repeat Tau Regulate the Dynamic Instability of Two Distinct Microtubule Subpopulations in Qualitatively Different Manners: Implications for Neurodegeneration. *Journal of Biological Chemistry*, 280(14):13520–13528, 2005.