1 Introduction

PhDCOMICS is very popular among energized (and sometimes frustrated) Ph.D. students. One of the highest rated comics is “How Grad School is just like Kindergarten,” which compares Ph.D. students with kids in kindergarten. Referring to myself as the kid, I feel so fortunate to have had Professor Jaime Carbonell as my mentor in the wonderland of research. He not only gave me guidance in conducting research, but also exemplified how to be a true scholar.

Jaime is a world-renowned researcher in machine learning and natural language processing. My thesis work pushed the frontiers of these two areas to a third exciting area: deciphering biological languages using machine learning and human language techniques. Following the comics theme, I want to relate my experience with Jaime in “kindergarten.”

2 All Day Napping is Acceptable

Jaime always shielded me (and most of his students) from the difficulties of getting funding. So we, as students, could enjoy the freedom to do bluesky research and sometimes daydream about the coolest ideas (which usually turned out to be dumb ones). When I started working on my thesis project, Jaime told me to work on protein structure prediction via free energy minimization from physics. At that time I had no deep knowledge of biology or physics. Instead of working hard to learn those topics, I, like many naughty kids in kindergarten, wandered around (or maybe napped) trying to figure out if I could think of something cool and relatively easy to do. Later I settled down with the following research problem:

An important step in automatically inferring protein structures from amino-acid sequences is to identify the typical spatial arrangements of well-defined secondary structures, which are conserved over proteins in different organisms and/or from different species, i.e. structural motif. From computational perspective, we can represent a structural motif abstractly by a state sequence of its secondary structure components (such as \(\alpha\)-helix, \(\beta\)-sheet and coil) and the constraints between them (i.e. chemical bonding among secondary structure components). For example, one important motif in binding the ligands or substrates is the \(\beta\)-\(\alpha\)-\(\beta\) motif. It consists of three secondary structure components, including a \(\beta\)-strand, an \(\alpha\)-helix and another \(\beta\)-strand that forms hydrogen bonds with the first \(\beta\)-strand. Therefore we can abstract the motif as a state sequence “\(B_1 - A - B_2\)” and hydrogen bonding between the two \(\beta\)-strands as the constraints in the objective function. The goal of structural motif recognition is to predict whether the motif of interest exists in the testing protein sequence. Computationally, it can be achieved by segmenting and labeling the testing sequence against the motif template, i.e. the state sequence and the constraints.
3 There is Constant Adult Supervision

Jaime has a busy schedule, but he always tries to make time for his students (although, of course, we need to be smart about stalking the available slots and aggressive about grabbing them). Jaime and I used to meet regularly each week. Research, as its nature, sometimes progressed slowly (due to classes, paper rejections or fighting with one’s boyfriend). Jaime was always been patient and gives “adult supervision,” or advice, only as necessary. The way he expressed it would be by saying the following with a smile, “You did great, but probably we need to move forward.” As far as I remember, he said this to me only twice during my Ph.D. years. Luckily, I was able to move forward and formulate the hard biological problem into a hot machine learning topic:

From the machine learning perspective, protein structural motif recognition can be cast as a structured prediction problem. Structured prediction refers to the task in which the observed data are sequential or with other simple structures while the outputs actually involve complex structures. More specifically, in the task of protein structure prediction, we are given the observation of a sequence of amino acids, but the target outputs involves complex three-dimensional protein structures. By considering the constraints or associations between the outputs (beyond the i.i.d. assumption), known as the structured input output information, we are able to achieve a better prediction performance.

4 You Get FREE Pizza for Lunch

Jaime is extremely supportive in cultivating students, with pizzas and beyond. Carnegie Mellon has a famous machine learning seminar, called “Machine Learning Lunch.” It is held on Monday at noon, when hungry grad students can feed themselves with FREE pizza (or sandwiches) and food for thought. In the seminar, I met a lot of machine learning researchers and students. Many of them become good friends, collaborators, and thesis committee members. In addition to free pizza, I was also given the opportunity to travel to conferences and meet with researchers outside CMU, even if I did not have a paper or I was not the first author of the paper. (By the way, at that time 30% of ICML papers came from CMU. So we had a higher likelihood of running into someone from CMU than from any other single place at the conference). By collaborating with Professor John Lafferty, Eric Xing, Vanathi Gopalakrishnan and many others, I successfully landed the solutions to my thesis problem:

Following the graphical model approach for structured input output predictions, we proposed a series of new models for our task of structural motif recognition. These models can be seen as an extension of the conditional random fields (CRF) by joint modeling the constraints between the components either on one sequence or multiple sequences. The key questions is how can we better represent the structural information of a motif using the graphical model and given the foreseeable complexity of the model, how we can learn the parameters of the model and make inferences efficiently. Therefore we developed kernel CRF, segmentation CRF, and chain CRF to solve the problem and demonstrated their effectiveness on our application.

5 There Are No Grades, But There are Rejections

Every Ph.D. student has experienced paper rejections, despite having spent so much time and effort (sigh! Those sleepless nights...). Different from the productive Ph.D. students these days (who publish 3-4 papers per year), we usually worked on one paper each year. Therefore getting a paper rejected literally meant a denial of a whole year’s work (of course it’s not, but to a student it appears to be). The nature of my thesis problem, i.e., solving a challenging biology problem with advanced machine learning techniques, meant that it naturally took more effort, especially as we aimed to make contributions in both fields, and thus would endure more rejections if we were not well prepared. Jaime was always supportive during those tough days. He never asked me what went wrong, instead he usually chatted about some new research directions (or his chess games with Professor Raj Reddy). After I got out of my mad mood (usually one week after), he would sit together with me and examine the reviews. After several rejections, we started to find our way to successes and published the research results in top conferences in both computational biology and machine learning.
6 Crying For Your Advisor After Graduation is Normal

I can go on and on telling all the details that I experienced with Jaime in my long period as a Ph.D. student, but everything has an end. The same is true for my Ph.D. days. The day finally came when I defended my thesis, landed a job, and prepared for a new life after graduation. My last meal in Pittsburgh before leaving was a joint dinner with Jaime, my husband and his Ph.D. advisor, Professor Alex Hauptmann. I remember it was raining and everything felt so different. We had champagne to celebrate, but it was very hard for us to say goodbye to our advisors. Jaime kept his classic smile on his face and said “Well, good luck at IBM. We’ll miss you.”

Yes, I missed Jaime a lot after graduation, especially when I felt frustrated, puzzled, or hoped for someone to save me from the turmoils of work. However, I would prefer to say that I am very proud that I survived those challenges thanks to Jaime’s mentorship in this wonderful kindergarten. I feel lucky to have experienced my best days with Jaime, who has such a great mind.

Thank you, Jaime, and Happy Birthday!