

Undergraduate Internship Program Report – Spring 2015

Renato Negrinho

In this short report, I briefly describe my experiences during my three-month internship (from February to April 2015), hosted by Prof. Noah Smith at the Language Technologies Institute in the School of Computer Science at Carnegie Mellon University.

My work during that period focused on developing a multiword expression tagging model based on an adaptation of the latent conditional random field for the semi-supervised setting. We chose the web reviews domain as multiword expressions are very common there. As data for training the model, we considered a small hand labeled corpus and a large unlabeled corpus, namely the Yelp Academic Dataset. We discussed different ways of training the model. We also identified possible limitations of the model as first conceived, and developed some variations that tackle these perceived problems. Due to the recency of the work we built on, there was little code that we could reuse, which implied carrying out a significant design and coding effort to run all the experiments. As a consequence, some work remained to be done at the end of internship, but I will finish it before submitting a paper to a conference or writing a final technical report.

Overall, it is hard to overstate how rewarding I feel my experience at CMU has been. I was fully integrated in Prof. Noah Smith's group. I attended the group meetings and had weekly meetings with him, to discuss progress in the project and research in general. One of his senior PhD students, Waleed Ammar, was also involved, and we met with some regularity. I also got to know and interact with the other students in Prof. Noah Smith's group and met other computer science faculty and students at CMU.

The School of Computer Science is great. There is a large number of regular seminars series on very interesting topics as, for example, Machine Learning, Natural Language Processing, Statistics, and Computer Science Theory. There is also a work ethic at CMU that is second to no other place. People really do work very hard there. With such an intense environment, one can feel really that they have the critical mass necessary to conduct cutting edge research.

Being at CMU during that particular period also afforded me an experience which would be very hard to get otherwise. I had applied for PhD programs at US universities, and during the months of February and March, you get to hear the admission decisions and attend the open houses for the programs you have been admitted to. I am happy to announce that I have been admitted to some of the programs I have applied to. Being in the US during this period allowed to attend the open houses and talk to faculty, who are leaders in their fields of study, and to some of my fellow incoming PhD students, who hope to make relevant contributions. I believe I have acquired some important contacts as a consequence of these events. It was truly an unique experience. I am also happy to say that among the programs I have been admitted to, I chose to attend the Machine Learning one at CMU and therefore, I will be returning to Pittsburgh in the Fall to pursue my PhD studies.

Finally, I want to thank Prof. Noah Smith, for hosting me and for the many interesting research discussions. He was able to provide a stimulating and nurturing research environment within his group. I also want to dearly thank the CMU Portugal Program for the bold initiative that is the Undergraduate Internships Program. This program provides a much needed bridge between Portuguese academia and the Computer Science research powerhouse that is CMU. Without the support from the program, hardly any of the experiences described would have been possible. I also want to thank Dr. André Martins, Prof. Pedro Aguiar, and Prof. Mário Figueiredo for all the guidance and encouragement that they so kindly have provided for this internship at CMU and my PhD application.