

Cover letter Unbabel - Summer AI Research Internship

Unbabel is precisely the company I want to work for. Your company accepts the challenge of making state of the art research in machine translation work in real world applications. I find that inspiring. I have been following your company for some months now, and was delighted to learn of the internship opportunity offered by you this summer. The internship was brought to my attention by my thesis supervisor Wilker Aziz, who was so very kind to mention me to his academic collaborator, Andre Martins. This connection moreover cemented my belief that Unbabel has a unique intersection with the research community. I am convinced that I provide precisely the right skills to make this internship a success and hope that by the end of this letter you will think so too.

Should I mention all these courses? My attached CV shows that I have a diverse academic background. I consider this a major strength: moving between programs has taught me how to adapt quickly, work hard, and see the larger picture. I used the period in the bachelor of mathematics to transition into the more formal work I do now, and acquired a deep interest in mathematics in the process. In the master's program I developed an interest in computer science from a theoretical viewpoint, and followed courses on computational complexity, discrete mathematics, information theory and quantum computing. I rejoined with my interest in language when following courses on Natural Language Processing and found a home for my interest in mathematics in the courses on Machine Learning I completed.

Currently I am writing a master's thesis under supervision of Prof. Wilker Aziz, in which I investigate the question: What are effective ways of incorporating syntactic structure into neural language models? I study a class of neural language models that explicitly model the hierarchical syntactic structure in addition to the sequence of words. This class of models merges generative transition-based parsing with (recurrent) neural networks to parametrize the transition model.

This might be too detailed... These are fundamentally joint models, but can be evaluated as regular language models, modeling only words, by marginalizing over the latent syntactic structure. I focus on one model in particular: Recurrent Neural Network Grammars (RNNG) [Dyer et al. 2016], a model for which exact marginalization is intractable, but where importance sampling, using an external discriminative parser as proposal distribution, provides an effective approximate method.

This might be hard to understand so by itself... I take the RNNG and investigate the effect of using a globally trained chart-based parser as proposal model, investigate semi-supervised learning with unlabeled data, and finally I the model's syntactic robustness through their grammatical acceptability judgements. I also investigate in general the added value of making syntax a discrete latent

variable, by comparing the RNNG to a robust and competitive alternative: multitask learning of a regular neural language model with a syntactic side objective.

This research mixes linguistically informed questions, mathematically and computationally elegant solutions, and rigorous targeted evaluation. This is precisely the research that I enjoy and that I excel at. The code repository is currently private, but will be made public upon presentation of the thesis, or it can be shared upon request.

Besides the coursework that I completed, I learned a lot from working on individual projects. Some examples are the following, which can be seen on my Github page:

1. Graph-based and transition-based dependency parsing with different learning algorithms (with neural network features, and a structured perceptron on manually defined features).
2. Latent Dirichlet Allocation with different inference methods: collapsed Gibbs sampling; (Stochastic) Variational Inference; and Amortized Variational Inference, using an inference network.
3. A number of didactic projects, implemented with teaching in mind: CKY parsing with rules and probabilities estimated from a treebank; count based ngram language models with smoothing and neural ngrams models; and word embeddings (GloVe, SVD on a (weighted) cooccurrence matrix).

This might be redundant? I strive to make these projects well-organized, well-documented and self-contained. I increasingly learn how important such rigor is in promoting transparency and reproducibility of research. I believe that the code for my thesis is the culmination of this learning process.

This might be redundant? In these projects I make regular use of the great libraries that the Python ecosystem has to offer: Dynet and PyTorch, and earlier some Tensorflow, wherever I need neural networks; Cython and Numpy for speed and general numerical computation; multiprocessing and joblib whenever I see the opportunity for parallel computation; matplotlib, seaborn, and pandas wherever I want to inspect, manipulate and visualize data.

This is nice, but is it really relevant for THIS position? I had the amazing opportunity to teach a group of students at the master level in a course-long research project. The project had them implement state of the art NLP techniques from scratch, and write a report on their experimental findings - an experience that was a first for plenty of students. I supervised and taught almost entirely by myself; a bold initiative from my side that was rewarded with great satisfaction and immense experience. The student's appreciation of my efforts was evident in their consistent attendance to the weekly, non-compulsory, meetings, their impressive results, and the enthusiastic (anonymous) feedback I received afterwards. This experience has taught me that I have it in me to be a great teacher and would love to do more of it.

I hope this letter has conveyed to you my excitement for the open position, and that I have convinced you that I am precisely the applicant you are looking for. I look forward to hearing from you.

Kind regards, Daan van Stigt