Sepehr **Sameni**

SOFTWARE ENGINEER · MACHINE LEARNING ENTHUSIAST

Tehran, Iran

(+98) 9331123581 | ✉ Sepehr.Sameni@gmail.com | Separius | separius

"Be the change that you want to see in the world."

Microsoft Recruitment Team   *January 28, 2019*

MICROSOFT INC.

Job Application for AI Residency

Dear Recipient,

MyObjectives

In my opinion, human intelligence lies in our language. With the help of language, we are able to stand on the shoulders of giants, name new concepts and further develop them. So I think language is the most important aspect of an intelligent entity and mimicking humans is the most promising way to achieve this end, thus NLP(natural language processing) is crucial for AI and even though a lot of improvements have been achieved in the past years we are still not there.

I think that all the current methods do much more of memorizing rather than understanding the language(we can see this by querying state of the art SQUAD 2.0 models with out-of-domain questions). The hard thing about mid-level NLP tasks(like selecting the answer from a context given a question) is the lack of largely annotated corpora, therefore, we have switched to unsupervised methods like word2vec and BERT. Foregoing methods were a great step forward however the absence of something (more understanding-oriented methods?) is bothering as we are still using the memorization techniques to map BERT embeddings into our high-level targets.

In my opinion, we need something like capsule nets for NLP; an inherently generalizable method for understanding language usable in high-level tasks (like conversational agents). As language is ambiguous in general and it highly depends on one's beliefs, I believe generative models, by learning a none one to one mapping, are a great tool to tackle NLP and can help us achieve this goal.

How AI Residency Helps?

First and foremost being an AI resident is a great opportunity for me to work with other knowledgeable and ambitious people, ideally with different backgrounds. Sadly few people work in deep learning in Iran and my advisor (Mohammad Amin Sadeghi) is a computer vision expert, not an NLP one. I study in one of the best universities in Iran and we have a great undergrad curriculum but most of the intelligent ones leave the country as soon as they finish their undergrad studies as they find more worthwhile opportunities abroad, therefore, those who stay look for the lowest hanging fruits, nothing risky or innovative. Let me give you an example: when I learned about BERT, the first thing that came to my mind was to apply the masking technique on other tasks (like graph embedding or gene embedding) or what if we used a TCN(Wavenet) instead of the transformer. Even more general, now that we know transformers are great, let's improve the running time of them (from $\mathcal{O}(n^2)$ to something like $\mathcal{O}(n*k)$) but others (in my university) just rush into applying the same thing on let's say Arabic sentiment analysis which is not an extrapolation of the idea. I believe Microsoft is the right place for me to do ambitious things I am not able to do here for it has the right environment and people.

I'm also excited about the microsoft resources. back in our AI lab we only have one GTX-1080 and doing research in this fast era requires a lot more (for a lab of ten people). Sadly because of the US sanctions, we can't rent GPU and there aren't any GPU cloud providers in Iran. I also can't work in a big Iranian company and use their resources and not to anybody's surprise our industries, technological-wise, are far behind and it's hard to do real research and development here. Microsoft AI residency, on the other hand, flourishes new ideas and provides a lot of resources to do research (based on what I have read).

Last, but not least, I love sharing. I love the philosophy behind free software and I believe in open science. I can share my work with others if I join the program and this is truly important for me. And as a side note, I am a software engineer and I admire high-quality code (most of the research codes are terrible) and I'm sure that I can learn a thing or two from Microsoft's code standards and hopefully share what I will learn with my colleagues in Iran.

Past Experience

Given the current advancements of deep learning in the medical sciences, we decided to work on an early Alzheimer detection method, most of the current methods use PET or other costly brain measurements, but obtaining EEG is relatively cheap and thus we focused on EEG. To our surprise, there weren't any major publications on EEG and deep learning. This lack of interest is mainly due to the fact that there aren't any huge labeled datasets for EEG and we also couldn't annotate our own dataset because of the costs. So instead we focused on unsupervised methods to generate realistic EEG signals with the newly introduced EEG corpus called TUH-EEG.

Following the common procedure, I started with a simple model, a convolutional VAE (variational autoencoder) and the loss over time was promising but the results were not good, even to my non-expert eyes. So we switched to a Wavenet like network and it was unable to learn anything. after a lot of thinking and debugging, I realized that the input data is dirty and needs a lot more of cleansing (than basic preprocessing methods). Two non-ml people with domain knowledge joined the team and improved the data selection methods so we managed to fit our Wavenet network to the data and despite that, it failed on multi-channel EEG prediction (it could only model single channel EEG data which is not that useful). Further investigation revealed that the higher order interactions between channels are hard to model and we need a stronger Wavenet and naively increasing the number of parameters lead into overfitting so we need more date in order for the larger Wavenet to learn them which is not feasible. So once again we switched our method, this time to GAN(generative adversarial networks). I tried a lot of architectures and even though it was easier for the GAN to learn the multi-channel problem, it was harder to model really long sequences with high frequencies with our GAN model. After many modifications, I finally realized that generating EEGs can be easily modeled as a progressive method. I changed our architecture to a progressively growing one(PG-GAN) and it worked flawlessly.

As mentioned previously, there aren't any huge labeled EEG corpora out there and so we can't use Inception Score or FID which are widely used in image generation models, to evaluate our results. So once again I relied on feature engineering and used domain knowledge to extract non-linear features from real samples and generated samples and train an SVM over them to evaluate the goodness of our results.

Right now we are working to make our models conditional and doing so, without labels, is hard. So we are trying to do an information-maximization method but it's too early to say (if it will lead to the right place)

Sincerely,

Sepehr Sameni