

Personal Statement

I am enthusiastically applying for the Computer Science Doctoral Program at Northeastern University, driven by a passion for computational linguistics and a commitment to preserving linguistic diversity and cultures.

Academic Background

My academic journey in computer science at Sharif University of Technology, complemented by an ongoing Masters in Artificial Intelligence at Amirkabir University of Technology, has laid a robust foundation in natural language understanding, image and text processing, information retrieval, machine learning, big data engineering, scientific computing, mathematics, and theoretical computer science.

Ranking in the top 0.3% on the national university entrance exams and graduating from the most competitive university for STEM in Iran, alongside my history in the chemical engineering major (ranking 3rd of 97 in the first two years), followed by a transfer to the mathematics department, demonstrates my consistent resilience. Additionally, my experience in systems biology research, NLP projects, and teacher-assistantship for various courses showcase my adaptability.

Research Interests

My current research on neural information extraction in Persian and my projects reflect a deep engagement with language and technology. The potential of NLP for social impact, particularly in understanding and preserving low-resource languages, means a lot to me. I believe language varieties, such as different social media jargon and historical texts, can be treated similarly to low-resource languages, and doing so can unlock many research opportunities. These variations contain unique words and cultural nuances and provide deeper context towards their corresponding community.

Projects

Gilaki, an endangered local language, has been the topic of my self-driven research for the past two years. A comprehensive linguistic presentation, developing NLP tools and preprocessing texts, gathering a corpus and resources, a comparison with Kurdish variations and other ethnicities' languages, studying different dialects of Gilaki and Kurdish, and interviewing natives are some contributions I made.

I have always welcomed learning new skills and programming languages and tackling real-life self-driven projects, either as simple as an online whiteboard or as intimidating as online video telephony during pandemics, adapting distributed architecture and micro-services. For the latter, I looked into different network protocols to verify the possibility of using distributed data structures and peer-to-peer network architecture. I also took a course on big data engineering,

gaining a deeper understanding. Although the project was not finished, it provided insight and broadened my perspective.

It was not until the last year of my bachelor's degree that I decided to pursue natural language research as my career. This has led me to enter a Master's program this year where I can rapidly learn language and speech technologies and deep learning concepts while honing my research skills. In my current research, I am addressing the sub-optimal relation extraction performance of large language models and their limited understanding of Persian, a seemingly under-represented language in their training. To improve this, I am studying state-of-the-art techniques such as parameter-efficient fine-tuning and alignment of model encodings with other multilingual language models. I will leverage them to fine-tune models for under-represented languages and with low computing resources.

Why Northeastern

I am eager to contribute to and grow within Northeastern's community, particularly inspired by the works of Professors Mai ElSherief, David Smith, and Malihe Alikhani. Their research is highly enlightening and intersects significantly with both my values and my interests in social computing and linguistic analysis.

Coming from an oppressed region marked by the diversity of languages and ethnicities and the bitterness of political strife within it, I have seen first-hand the relentlessness against minorities, the increase of division and violence between groups, and the isolation of ones in need of help. Hence, I have the incentive to do impactful research on social well-being of different ethnicities and to mitigate hurtful behaviors. One approach is to identify marginalized groups by gathering data from key tweets mentioning them along with their comments. The users involved can be related to that group through inclusion, support, or opposition. The resulting distantly supervised data can be further used in research and practice, such as developing culture-aware language models.

At the same time, the works of Professor David Smith perfectly align with my passion for linguistic analysis. I am very enthusiastic about languages, observed cultures, and their hidden contexts. In addition, my familiarity with Persian literature and Arabic allows me to integrate deeper insights into my research. I would like to build upon his work by identifying new and creative ways to further detect language and cultural shifts and differences and understand the implications of such changes. As a future researcher, I plan to aid linguists by building tools for corpus collection and analysis and enhancing language models for documenting, learning, and using lower-resource languages.

Conclusion

The opportunity to study at Northeastern University, with its vibrant and diverse cultural landscape and access to cutting-edge research, is priceless to me. Located in Boston, a city rich in history and a center of academic excellence, Northeastern offers numerous opportunities

for me. The university's strong connections with other prestigious institutions, its access to conferences and internships, and its leading role in technological research provide an unparalleled environment for advancing my studies. Additionally, I have dear friends I hope to reunite with in Boston.

This program is pivotal to my dream of becoming a researcher, developing technologies for computational humanities, and multilingual language understanding. My tinkering nature, alongside my desire to ask, explore, learn, and make change are the reasons I can't wait to become a full-fledged researcher and an advisor for those who share my passion for science.

I believe my solid background, ambition, and resilience, combined with a strong incentive to make social change, make me an excellent candidate for this position, and with the emergence of larger language models, many more horizons are yet left unventured. I am excited about being part of your vibrant, interdisciplinary community.

In summary, this program represents the ideal environment to advance my skills, pursue my research interests, and prepare for a career that bridges the gaps between linguists, sociologists, and computer scientists.