Statement of Purpose

Zhiyu Yuan
yzy18@tsinghua.org.cn

My research interests lie in applying AI technologies to address real-world challenges, with a focus on enhancing machine learning systems and their applications, such as recommendation systems. My fascination with AI and its transformative potential has driven me to explore machine learning systems in both academic and industry settings. I am applying for the PhD program at University of California, San Diego, as it presents an unparalleled opportunity to collaborate with experts, further my research, and lay the groundwork for a fulfilling career in AI. Now, I am excited to share how I first encountered this field, the work I've done so far, and my plans for the future.

Research Experience

The Start of the AI Research My interest in machine learning began during the 2017 National College IoT Design Competition, where I led a team to develop an intelligent mat for office workers. The mat tracked working hours, detected sitting posture, and identified users via pressure sensor data. Although I had limited knowledge at the start, the experience ignited my curiosity. I taught myself the basics of machine learning and applied Support Vector Machines (SVM) to solve the challenge. Despite working with a small dataset of only three participants, the experience proved transformative. Our project won a national first prize, and I co-authored a paper [1] under Prof. Haojun Ai, which was later accepted at ICPR 2018. This success not only marked the beginning of my academic journey but also sparked my passion for AI, motivating me to explore the field further.

A Leap Forward at Deecamp Building on the foundation I developed during the IoT competition, I sought to deepen my understanding of AI by attending Deecamp the summer after graduation. I was inspired by lectures from renowned experts like John Hopcroft and Kai-Fu Lee, whose insights expanded my understanding of machine learning and deep learning. Motivated by the potential of AI, I joined a team focused on autonomous driving, a field that seemed both challenging and exciting. Over the course of one intense month, we developed an autonomous driving system from scratch, integrating modules such as high-precision mapping, object detection and segmentation, and path planning. Our project won the Best Application Award, and this experience was pivotal in transforming my theoretical understanding of AI into practical, hands-on skills. It solidified my confidence in applying AI to real-world challenges and further fueled my desire to pursue AI research.

Deepening Expertise in AI Following my enriching experience at Deecamp, I sought to formalize my AI expertise during my graduate studies at Tsinghua University. There, I fo-

cused on building a strong foundation in AI and its applications. I took advanced courses in Deep Learning and Artificial Neural Networks, earning A- and B+ grades, which reinforced my core knowledge in these areas. I also sought to apply my learning in practical settings, participating in a Kaggle competition on toxicity classification. This allowed me to apply deep learning techniques in the field of natural language processing, where my team ranked in the top 1% historically. This experience deepened my problem-solving skills and solidified my commitment to advancing in the AI field.

Exploring AI in Industry With a solid academic foundation in AI, I was eager to see how these concepts were applied at scale in industry. This led me to an internship at Kuaishou, where I contributed to the User-dependent Gating (UDG) project [2], which aimed to improve recommendation system precision by refining user-dependent models. The experience gave me invaluable insights into the challenges of optimizing AI systems in real-world applications. After my internship, I led the Hydrus project [3], focusing on balancing system resources with user experience. I developed a system that dynamically allocated resources to optimize performance while maintaining a seamless user experience. This system was successfully deployed at scale, benefiting millions of users. The results were published at SIGIR 2023, further cementing my expertise in machine learning systems and motivating me to continue pursuing impactful research in AI.

Future Vision

I aspire to broaden my experience in applying machine learning to a wide range of real-world applications. During my PhD, I aim to focus on inference optimization, particularly in light of the substantial resource consumption of large language models. The research conducted by Professors Hao Zhang and Yiying Zhang at University of California, San Diego aligns perfectly with my interests. I am particularly drawn to their work on efficient machine learning system, which I believe can provide the ideal environment to pursue this research. This will enable me to develop innovative solutions that have a meaningful impact on both industrial applications and society.

References

[1] H. Ai, L. Zhang, Z. Yuan, et al. "iCushion: A pressure map algorithm for high accuracy human identification". In: 2018 24th International Conference on Pattern Recognition (ICPR). IEEE, 2018, pp. 3483–3488.

[2] Yuan Z, Ren K, Zhang C, et al. Personalized Recommendation Systems with User-dependent Gating. unpublished manuscript. GitHub. URL: `https://yuanzhiyu.github.io/project/udg/udg.pdf`.

[3] Yuan Z, Ren K, Wang G, et al. "Hydrus: Improving Personalized Quality of Experience in Short-form Video Services". In: Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2023.