FAKULTA INFORMAČNÍCH TECHNOLOGIÍ ČVUT V PRAZE

Review statement on doctoral thesis

Ph.D. student: **Ing. MATEJ ČIEF**
Thesis title: OFF-POLICY EVALUATION AND LEARNING IN ADAPTIVE SYSTEMS
Supervisor: doc. Ing. MICHAL KOMPAN, Ph.D.

The full dissertation expands significantly on the author's preliminary investigations into off-policy evaluation (OPE) and reinforcement learning from human feedback (RLHF), particularly addressing challenges in large action spaces and reward hacking in language models. Matej Čief demonstrates extensive progress, enriching the work with deeper theoretical insights, comprehensive technical detail, and rigorous experimentation, notably through the implementation of embedding-based methodologies and adaptive uncertainty-aware frameworks, an especially important research direction in my opinion.

The theoretical foundations of the dissertation are solid, with a clear and nuanced understanding of crucial concepts in contextual bandits, OPE, and RLHF. The author's literature review is thorough and current, clearly illustrating awareness and integration of recent advances, thereby underscoring diligent scholarship and depth of technical expertise.

A significant strength is the innovative approach presented for learning action embeddings to improve OPE in large action spaces, particularly valuable given its successful empirical validation on synthetic and real-world datasets. This method addresses critical gaps in existing methodologies, enhancing practical applicability and robustness, especially for scenarios where predefined embeddings are unavailable or insufficient.

Equally commendable is the novel adaptive method developed to mitigate reward hacking in RLHF. The author's application of Products of Experts (PoE) to adaptively balance reward model uncertainty and regularization clearly demonstrates innovation and technical sophistication, effectively addressing a pervasive and challenging problem within the field. The empirical results convincingly validate the method's efficacy, showcasing its potential for broad applicability.

The dissertation is supported by multiple high-quality publications appearing in competitive, indexed venues. Impressively, these recent publications have already gathered 10 high-quality citations, highlighting their impact and relevance within the research community. The student's persistence in overcoming initial setbacks and rejections, now aspiring to publish in A* conferences, exemplifies admirable resilience and ambition—truly a commendable Slovakian academic success story. It is also noteworthy that the student successfully integrated into prestigious research environments such as Amazon internships, effectively collaborating with top researchers to further enhance the quality of his contributions.

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ ČVUT V PRAZE

One area for further improvement is clarifying the author's individual contributions versus those of collaborators, mentors, and colleagues involved in various stages of research. While the author appropriately acknowledges mentorship and collaborative guidance, explicitly delineating individual roles could provide clearer insights into personal scholarly advancements and responsibilities.

Additionally, it would be interesting to explore further the confidence prediction methods applied in this research. Could approaches like Monte Carlo Dropout provide effective estimates of confidence bounds or robustness for uncertainty estimation? Have you experimented with or considered alternative methods to further enhance the reliability and interpretability of confidence measures?

Additionally, I appreciate that Matej publishes and openly shares his research results (e.g., https://github.com/amazon-science/ope-learn-action-embeddings), significantly benefiting reproducibility. Given the fundamental and niche nature of this research, greater marketing efforts, perhaps leveraging platforms such as Amazon's promotional capabilities, could attract additional contributors and maximize its potential.

In conclusion, the dissertation constitutes substantial original research with both theoretical rigor and significant practical contributions to the fields of reinforcement learning and recommendation systems. Matej Čief has demonstrated exceptional capabilities in independent research and interdisciplinary collaboration, and I recommend acceptance of this dissertation by the evaluation committee.

doc. Ing. Pavel Kordík

Podepsal: doc. Ing. Pavel Kordík, Ph.D.

Datum: 2025-04-07 12:54:45 CEST

doc. Ing. Pavel Kordík, Ph.D.

Prague, Apr 7th. 2025