13 January 2025 ·
Qeios

Peer Review

Review of: "Fine-Grained Alignment in Vision-and-Language Navigation through Bayesian Optimization"

Yufeng Yue[1]

1. Beijing Institute of Technology, China

The paper presents a well-structured and innovative contribution, addressing the fine-grained alignment issue in Visual-Language Navigation (VLN) by proposing a Bayesian optimization-based adversarial training framework. The proposed method not only has a solid theoretical foundation but also demonstrates significant improvements in generating fine-grained visual negative samples. Experimental results on the R2R and REVERIE datasets validate the effectiveness and wide applicability of the framework. Below are the strengths of the paper and some minor suggestions for improvement:

Strengths:

1. The combination of Bayesian optimization and fine-grained negative sample sampling provides an innovative solution, which has been well validated in terms of enhancing both discriminative and generative performance in VLN tasks.

2. The experimental design is comprehensive, including multiple baselines, ablation studies, and performance comparisons with existing methods, which demonstrate the effectiveness of the framework.

3. The paper clearly highlights its contributions, particularly in improving visual-language alignment and performance by generating fine-grained visual negative samples.

4. The availability of code and pre-trained models ensures high reproducibility, which will be helpful for further research based on this work.

Suggestions for Improvement:

1. A more detailed discussion on the computational complexity introduced by Bayesian optimization, particularly its impact on training time, and an evaluation of its feasibility in real-time applications

 doi.org/10.32388/0LP72J

would be beneficial.

2. It would be helpful to expand the discussion on how the framework can be applied to other visual-language tasks, such as Visual Question Answering (VQA) or image captioning, to showcase the broad applicability of the method.

3. More comparison figures for path selection could be added to illustrate the actual impact of fine-grained versus coarse-grained negative samples on navigation path selection.

4. A visual representation of the Bayesian optimization process, such as an optimization evolution graph, could help readers intuitively understand how the optimization selects the most challenging negative samples.

5. The formula for generating negative samples in Bayesian optimization could be further explained (e.g., Equation 4). Additionally, a simple derivation of the optimization objective function (Equation 5) would clarify how Bayesian optimization impacts training. Also, some formulas are not numbered.

6. A deeper analysis of the impact of strategies like delayed updates and out-of-domain surrogate frames would be valuable, potentially through graphical demonstrations of performance variations under different strategies, to further elucidate the role of these design decisions.

Declarations

Potential competing interests: No potential competing interests to declare.

doi.org/10.32388/0LP72J