Improving Generative Visual Dialog by Answering Diverse Questions

**نویسنده اول: Vishvak Murahari**

**سال انتشار: 2019**

**محل چاپ: EMNLP**

**لینک به صفحه دانلود مقاله: https://arxiv.org/abs/1909.10470**

## در یک جمله بنویسید مقاله چه کار کرده

This article has proposed an approach to improve the training of generative Visual Dialog models by encouraging diverse questioning and reducing repetition during self-talk interactions, resulting in more varied, coherent, and relevant dialogues about images.

## چکیده مقاله

Prior work on training generative Visual Dialog models with reinforcement learning (Das et al., 2017b) has explored a Q-BOT-ABOT image-guessing game and shown that this ‘self-talk’ approach can lead to improved performance at the downstream dialog-conditioned image-guessing task. However, this improvement saturates and starts degrading after a few rounds of interaction, and does not lead to a better Visual Dialog model. We find that this is due in part to repeated interactions between Q-BOT and A-BOT during self-talk, which are not informative with respect to the image. To improve this, we devise a simple auxiliary objective that incentivizes Q-BOT to ask diverse questions, thus reducing repetitions and in turn enabling A-BOT to explore a larger state space during RL i.e. be exposed to more visual concepts to talk about, and varied questions to answer. We evaluate our approach via a host of automatic metrics and human studies, and demonstrate that it leads to better dialog, i.e. dialog that is more diverse (i.e. less repetitive), consistent (i.e. has fewer conflicting exchanges), fluent (i.e. more human-like), and detailed, while still being comparably image-relevant as prior work and ablations. Our code is publicly available at github.com/vmurahari3/visdial-diversity.

## کارای قبلی که مقاله ذکر کرده چی بودند (background literature, previous work)

The article refers to the following previous works:

Visual Dialog: Prior research by Das et al. (2017a) and de Vries et al. (2017) in the field of visual dialog.

Diversity in Text-Only Dialog: Zhang et al. (2018) and Li et al. (2016a,b) explored methods for modeling diversity in text-only dialogues.

Conditional Variational Autoencoders: Massiceti et al. (2018) used conditional variational autoencoders for self-talk in visual dialog, focusing on diversity measurement.

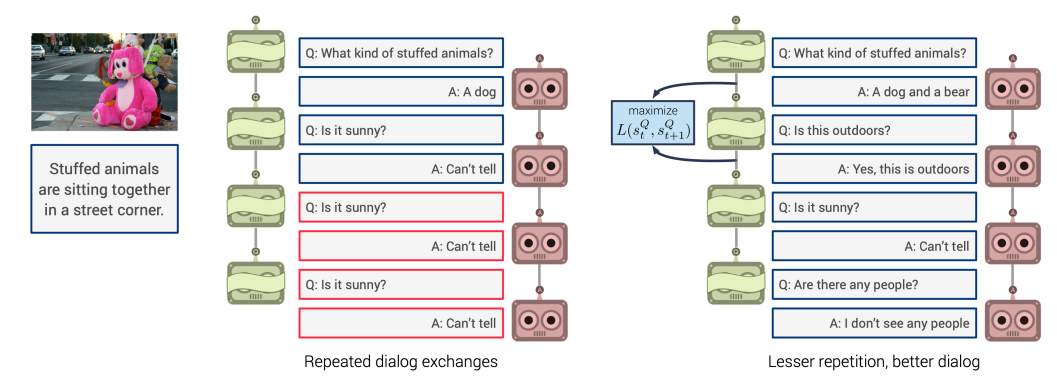
Constraints and Diversity: Zhang et al. (2018), Gao et al. (2019), and Li et al. (2016a) studied approaches to generate diverse responses in dialog, either through adversarial information maximization, modeling diversity and relevance, or maximum mutual information criterion. The article distinguishes itself by focusing on diversity across entire dialogues and reducing repetitions

## به طور خلاصه در یک پاراگراف بگید که مقاله، چالش، ایراد یا ضعف کارهای قبلی رو چگونه ارزیابی کرده

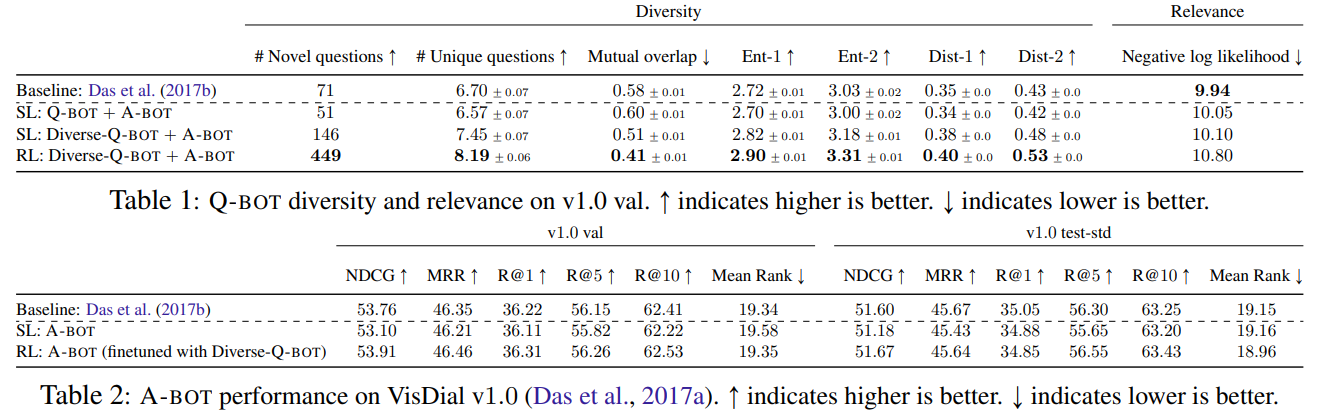
Repetitive Interactions: The prior work involving self-talk interactions between AI agents (Q-BOT and A-BOT) resulted in repeated interactions that weren't informative for understanding the image, leading to a plateau and eventual degradation in performance.

## ایده مقاله برای حل چالشها، ایرادها یا ضعف های پاراگراف قبل چی بوده (خلاصه)

مدل Q-BOT رو به گونه ای طراحی میکنن، که سوالهای تکراری نپرسه و بتونه سوالات متنوع رو بپرسه.



## مقاله تا چه حد تونسته با روش پیشنهادی خودش به اون چالشها، ایرادها و ضعف ها پاسخ بده (metrics,evaluations,results)



از نظر تنوع سوالات بالاترین معیارها رو کسب کرده و از لحاظ Performance تقریبا با baseline model یکسان هست.

## شما برای حل چالشهای گفته شده چه ایده ای دارین؟ (خلاصه)

ایده ای ندارم چون خیلی از مفاهیمش رو متوجه نمیشم.

## شما برای بهبود این مقاله چه ایده ای دارین؟ (خلاصه)

استفاده از چندین Q-BOT به طور همزمان