Proceedings of the Canadian Conference on Artificial Intelligence

Dataset Augmentation Using Back-Translation to Improve Early Stage Dialog Systems

Marc Queudot, Louis Marceau, Raouf Moncef Belbahar, Eric Charton, Marie-Jean Meurs

Published on: May 27, 2022

ABSTRACT

As dialog systems are increasingly used, a major challenge for building new ones is the lack of annotated training data. The necessary data collection and annotation efforts are laborious and time-consuming. A potential solution is to augment initial seed data by automatically paraphrasing existing samples. In this paper, we propose a novel data-efficient approach towards this goal. Our method can kick-start a dialog system with minimum human effort while delivering a performance strong enough to allow real-world usage. We ran experiments using Neural Machine Translation on two open corpora. On both of them, the proposed approach improved the generalization capabilities of the model. Our results suggest that paraphrase generation techniques could be used as-is to provide a boost in performance to dialog systems in an early phase.

Article ID: 2022S06

Month: May

Year: 2022

Address: Online

Venue: Canadian Conference on Artificial Intelligence

Publisher: **Canadian Artificial Intelligence Association**

URL: https://caiac.pubpub.org/pub/a7jrjex2