Continual Learning of Long Topic Sequences in Neural Information Retrieval - abstract

Thomas Gerald¹, Laure Soulier¹

¹Sorbonne Université, CNRS, ISIR, F-75005 Paris, France

Abstract

In information retrieval (IR) systems, trends and users' interests may change over time, altering either the distribution of requests or contents to be recommended. Since neural ranking approaches heavily depend on the training data, it is crucial to understand the transfer capacity of recent IR approaches to address new domains in the long term. In this paper, we first propose a dataset based upon the MSMarco corpus aiming at modeling a long stream of topics as well as IR property-driven controlled settings. We then in-depth analyze the ability of recent neural IR models while continually learning those streams. Our empirical study highlights in which particular cases catastrophic forgetting occurs (e.g., level of similarity between tasks, peculiarities on text length, and ways of learning models) to provide directions in terms of model design. The integral version of the paper has been published at ECIR 2022.

Continual learning, information retrieval, Neural ranking models