

▶ 数据简介

对于推荐系统社区来说，数据在其中扮演着一个重要的角色，它能够用于评估模型，一系列优秀的数据集。然而现有的数据集要么数据规模较小，要么数据集内容比较单一。导致难以充分测试出学发布一个大规模的多场景数据集，来推动推荐系统社区的发展。

Tenrec是一个大规模多场景的真实世界数据集，从腾讯的两个内容平台应用程序（包含视频推荐与图

下载

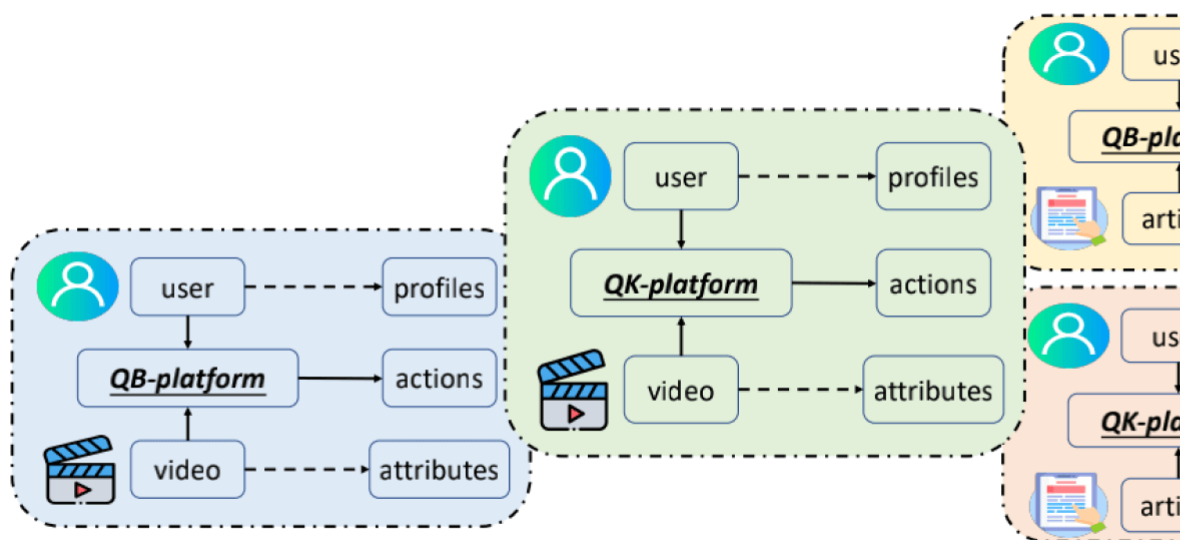


Figure 1: Data sources of Tenrec.

该数据集包含四个场景，我们随机从QK-video的数据库中抽取5百万的用户，抽取时选择点击行为我们的用户反馈（包括点击、分享、喜欢和关注等）、真实负样本（有曝光但是没有用户行为）。在G1千万的喜欢行为，1百万的分享行为和86万的关注行为。除此之外，数据集还包含经过匿名化处理的特征。同样的，我们在另外三个场景QK-article, QB-video和QB-article采用相似的策略来抽取数据。