

基于图卷积注意力网络的视频摘要方法

李平^{1,2}, 唐超¹, 徐向华¹

¹杭州电子科技大学计算机学院, 中国杭州市, 310018

²南京大学计算机软件新技术国家重点实验室, 中国南京市, 210023

摘要: 视频摘要已成为生成浓缩简洁视频的一种基础技术, 该技术有利于管理和浏览大规模视频数据。已有方法并未充分考虑各视频帧之间的局部和全局关系, 导致摘要性能下降。提出一种基于图卷积注意力网络 (graph convolutional attention network, GCAN) 的视频摘要方法。GCAN由嵌入学习和上下文融合两部分组成, 其中嵌入学习包括时序分支和图分支。具体而言, GCAN使用空洞时序卷积对局部线索和时序自注意力建模, 能有效利用各视频帧的全局线索; 同时利用多层图卷积网络学习图嵌入, 反映视频帧样本的本征结构。上下文融合部分将时序分支和图分支的输出信息流合并形成视频帧的上下文表示, 然后计算其重要性得分, 据此选择具有代表性的帧, 生成视频摘要。在两个基准数据集 SumMe和TVSum上的实验结果表明, 相比其他多种先进方法, GCAN方法在3种不同评测环境下取得更优越的性能。

关键词: 时序学习; 自注意力机制; 图卷积网络; 上下文融合; 视频摘要

<https://doi.org/10.1631/FITEE.2000429>