

摘 要

随着互联网技术的快速发展，图像和视频等多媒体数据呈现出爆炸式的增长趋势，这也对相应的信息处理技术提出了新的挑战。一方面，受限于计算机的处理能力，人们需要将有限的计算资源分配至重要视觉信息进行重点分析和处理；另一方面，人们希望使用计算机获得的自动分析结果能够符合人类认知。针对这两方面的需求，本文将主要研究如何通过机器学习的方法进行视觉显著计算，以实现符合人类认知的重要视觉信息的自动预测、定位和挖掘。本文的主要创新点包括：

第一，提出了一种基于概率多任务学习的视觉显著计算方法，用于解决自底向上和自顶向下因素的融合问题。该方法将多任务学习算法引入视觉显著建模过程中，这在国内外相关研究工作中尚属首例。该方法通过一个概率框架同时引入了自底向上和自顶向下因素的影响。该框架不仅使用多尺度小波分解来模拟视知觉系统中的底层过程，而且通过学习算法来建模视知觉系统的高层过程中对输入信号的偏置。在该框架下，本文通过多任务学习算法来同时优化在不同场景上的模型以及模型融合策略。通过在多种数据集上的对比实验，该方法能够取得较好的视觉显著预测效果，并具有较高的健壮性。

第二，提出了一种基于代价敏感配对排序学习的视觉显著计算方法，用于解决视频眼动数据集的稀疏标注问题。该方法在国内外相关研究工作中首次将视觉显著计算建模为排序学习问题，具有鲜明特色。该方法避免了对可信正例和负例的直接选择，而是通过一种代价敏感的排序学习框架来直接引入稀疏正例和未标注数据的影响。实验证明，该排序学习框架能够同时引入局部视觉特性和“显著目标—干扰物”之间关系的影响，使得该方法能够更好地适用于只有稀疏标注的视频眼动数据集。

第三，提出了一种基于多任务排序学习的视觉显著计算方法，用于解决视觉显著计算时的场景自适应问题。该方法将视觉显著计算建模为多任务配对排序学习问题，对不同的场景类别构建不同的视觉显著模型。在模型构建过程中，该方法不仅能够同时学习出多个适用于不同场景类别的视觉显著模型，还通过模型间的信息共享提高了每个模型的泛化能力。在视频眼动数据集上的大量实验证明，该方法在多种场景下均能够取得很好的视觉显著计算效果。

第四，合作提出了一种基于互补显著度图的显著对象提取方法，并在此基础上开发了基于视觉显著对象提取的视频广告关联系统。该广告系统包括广告拉取和推送两个子模块，分别通过简单用户交互和互补显著度图来提取视频中的兴趣/显著对象。此后，这些兴趣/显著对象将与用户兴趣结合起来，以一种低干扰的方式，为用户提供内容相关的个性化广告服务。目前，该系统即将被华为公司集成到其下一代智能流媒体服务产品中。

综上所述，论文针对基于学习的视觉显著计算所面临的三个重要问题进行了研究，同时对基于视觉显著计算的对象提取技术及其在视频广告关联中的应用作出了初步的探索。该论文在国际上首创性地总结了将机器学习方法引入视觉显著计算的必要性，并通过大量实验展示了通过机器学习的方法进行视觉显著计算的可行性和有效性。此外，本论文的研究工作为在视觉显著计算领域开展更为广泛深入的研究奠定了基础。

关键词：视觉显著；多任务学习；排序学习；场景自适应模型；视觉显著对象提取；视频广告关联