

In multimedia annotation, due to the time constraints and the tediousness of manual tagging, it is quite common to utilize both tagged and untagged data to improve the performance of supervised learning when only limited tagged training data are available. This is often done by adding a geometry-based regularization term in the objective function of a supervised learning model. In this case, a similarity graph is indispensable to exploit the geometrical relationships among the training data points, and the graph construction scheme essentially determines the performance of these graph-based learning algorithms. However, most of the existing works construct the graph empirically and are usually based on a single feature without using the label information. In this paper, we propose a semi-supervised annotation approach by learning an optimized graph (OGL) from multi-cues (i.e., partial tags and multiple features), which can more accurately embed the relationships among the data points. Since OGL is a transductive method and cannot deal with novel data points, we further extend our model to address the out-of-sample issue. Extensive experiments on image and video annotation show the consistent superiority of OGL over the state-of-the-art methods.