

**Multi-instance Learning with Discriminative Bag Mapping**

**Abstract:**

Multi-instance learning (MIL) is a useful tool for tackling labeling ambiguity in learning because it allows a bag of instances to share one label. Bag mapping transforms a bag into a single instance in a new space via instance selection and has drawn significant attention recently. To date, most existing work is based on the original space, using all instances for bag mapping, and the selected instances are not directly tied to an MIL objective. As a result, guaranteeing the distinguishing capacity of the selected instances in the new bag mapping space is difficult. In this paper, we propose a discriminative mapping approach for multi-instance learning (MILDM) that aims to identify the best instances to directly distinguish bags in the new mapping space. Accordingly, each instance bag can be mapped using the selected instances to a new feature space, and hence any generic learning algorithm, such as an instance-based learning algorithm, can be used to derive learning models for multi-instance classification. Experiments and comparisons on eight different types of real-world learning tasks (including 14 data sets) demonstrate that MILDM outperforms the state-of-the-art bag mapping multi-instance learning approaches. Results also confirm that MILDM achieves balanced performance between runtime efficiency and classification effectiveness.

**Existing System:**

Existing MIL solutions can be roughly divided into two categories: (a) updating a generic learning algorithm to tackle label ambiguity problems, or (b) developing a learning paradigm specifically for multiple instance learning. However, the performance of the above methods deteriorates when there are a large number of instances in a bag. Using content-based image classification, again, as an example, the total number of the instances in a bag could be extremely large if the image containes many regions. However, in reality, different regions/instances in a bag may make different contributions to image classification and the more informative the instances, the more information can be provided to learning tasks. In this scenario, selecting the most informative instances in each bag becomes a challenging problem for MIL.

**Proposed System:**

We propose a direct discriminative mapping approach for multi-instance learning (MILDM) that aims to identify the instances that will make the bags maximally distinguishable in the new mapping space, as shown in Figure 5. Experiments and comparisons on eight different types of real-world learning tasks (including drug activity prediction, content-based image classification, train bound challenge, mutagenicity prediction, scientific publication retrieval, online product evaluation, newsgroup categorization, and web index recommendation) confirm the effectiveness of the proposed design. The contributions of this paper are threefold.