## 3方法

在本节中,我们提出了EMLC(暂定名)方法的细节。首先介绍了如何基于Mlknn将多标签数据转换为单标签数据，然后针对新构造的数据提出了一种集成学习方法，进一步提高了算法对于正类的预测性能。

**3.1 EMLC算法**

为了提高正类的预测性能，我们提出了一种EMLC算法，该算法以MLknn的集成学习算法扩展而来。

对于多标签分类问题，我们使用采用传统的问题转换方法将多标签数据集转换为单标签数据集。算法伪代码在算法1中进行了描述。具体来说，构建新的单标签数据集一共有四步：1）将数据集划分为训练集和测试集。 2）在训练集上进行十次有放回的随机抽样，生成十份与原训练集同样大小的子集，再用Mlknn根据训练子集构建十个不同的基分类器。3）使用这些基分类器对预先划分好的训练集进行预测并得到每个标签的概率值，保存为一个新数据集，并且对测试集进行同样的操作。这样将每个标签在不同基分类器预测出的概率值组合成单标签数据集的特征向量。4）将原数据集的类标签插入到新数据集的最后一列构造出新的训练集和测试集。

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Algorithm 1.EMLC** | | | | | | | |
| **Input:** A Mutilabel data set D | | | | | | | |
| **Output:** two new single label dataset | | | | | | | |
| 1 |  |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |  |
| 4 |  |  |  |  |  |  |  |