标记分布学习作为一种新型的机器学习范式近年来正逐渐成为机器学习领域的热点研究问题之一。Geng等人为了解决年龄估计问题中训练样本不足的问题，首次提出了标记分布的概念，并且基于标记分布的思想设计了相应算法 IIS—LLD(Improved Iterative Scaling-Learning from Label Distribution)，成功应用在了自动年龄估计问题上，相比以往的年龄估计算法，由于IIS-LLD算法在训练过程中可以利用更多的样本信息，因此取得了更好的效果。

**1.标签分布的形式化**

标记分布学习的目标是以一种更加自然的方式去标记一个示例x，并且为它的每一个可能的标记Y分配一个数值，这个 数值被用来表示标记Y描述示例X的程度。不失一般性，假设∈[0,1]，进一步假设标记集合是完整的，也就是说使用标记集合中的所有标记才能够完整地描述示例。形式上可表述为。数值被视作标记Y对示例X的描述度。对一个特定的示例，标记集合中所有标记的描述度构建了一个类似于概率分布的数据形式。所以，我们称之为标记分布。学习一个标记示例的标记分布的过程就被称为标记分布学习。

标记y对示例X的描述度可以通过条件概率的形式来表示，即。然后，标记分布学习的问题可以按以下方式进行形式化的定义： 用来定义输入空间，表示完整的标记集合。给定一个训练集合，其中是一个示例，是与示例关联的标记分布，标记分布学习的目标就是要从训练集S中学习一个条件概率密度函数，其中x∈**X**,y∈**Y**。

假定是一个参数模型，其中是参数向量。给定训练集合S， 标记分布学习的目标就是在给定示例情况下，能够找到参数向量，然后用它来生成一个标记分布并且使这个标记分布与真实的标记分布的分布之间的距离最小或相似度最大。，如果我们使用KL散度(Kullback-Leibler divergence) 作为概率分布之间的距离标准的话，那么最好的参数向量将由以下优化式决定：

(1)

**2.IIS-LLD算法**

我们假设是一个既依赖于示例x又依赖于标记y的特征函数。那么，在训练集中特征函数关于经验联合函数的期望如下所示：

(2)

在训练集中，特征函数关于条件概率模型和经验分布函数的期望如下所示：

(3)

条件概率模型的一个合理选择是在满足约束条件的情况下，它拥有最大的条件熵。这样一个模型(亦称为，最大熵模型)是指数函数的形式，即：

(4)

这里是规范化因子，在中是第k个模型参数。实际上，这些特征函数仅仅依赖于示例，而不依赖于类别标记。因此，公式(4)可以被重写为如下形式：

(5)

这里是一个与类别标记无关的特征函数。

将等式(5)带入等式(1)中并且注意到，这样我们可以得出关于的一下目标函数：

(6)

注意到，等式(6)是基于最大熵模型的，众所周知，基于缩放的IIS(Improved Iterative Scaling，IIS)是最大化最大熵模型的似然函数的一种具有代表性的算法，因此，我们可以使用类似于IIS的策略对等式(6)进行优化求解。IIS算法开始于任意一组初始化参数，然后在每一次迭代过程中，它都会将当前参数的估计值更新为， 这里最大化了似然函数下界的变化。然而，在这里，迭代过程需要被转移到一个新的目标函数中。而且，IIS算法所要求的特征函数的约束条件(这里即为),应该被去除掉用以保证在选取适合数据的特征提取器时有更好的自由度。

详细地讲，目标函数的值在相邻迭代步骤变化如下：

（7）

这里是的增量。在等式(7)中应用不等式得：

（8）

对等式(8)右边部分关于进行微分得到关于的耦合方程，直接去求解这些方程是困难的。

为了解耦在之间的相互作用，我们在这里应用了詹森不等式(Jensen’s inequality)，即，对于一个概率密度函数p(x)来说，

(9)

把等式(8)中的最后一项改写为如下式子：

(10)

这里，是的符号。由于可以被看作是一个概率密度函数，将詹森不等式的结论应用到等式(8)后可以得到如下结论：

(11)

我们将等式(11)右边的部分定义为，它是的一个下界。将关于的导数置为零，即：

(12)

在等式(12)中是单独出现的，因此可以通过非线性方程求解方法(比如，高斯一牛顿法)将一个一个的求解出来。当被求解出来以后，我们就可以利用通过来更新模型参数直到算法收敛，最后通过最大熵模型的形式计算得到各个标记的描述度，算法具体的伪代码如下：

IIS-LLD算法：

1.算法输入：训练集，特征函数以及收敛阈值；

2.算法输出：

3.初始化模型参数向量；

4.：

5.重复如下计算过程,

Ⅰ.

Ⅱ.通过求解等式(12)得到；

Ⅲ.：

6.直至

7.

它以一种直接的方式解决了等式(1)中所描述的标记分布学习问题。