### 2.2.3 最大熵建模

最大熵建模是将各种不同来源的信息的知识聚集到一个统一的框架下。实际上所有的分类问题空间都可以表述成特征形式。这些特征可能很复杂，但是必须把和分类问题相关的先验知识描述出来。每个特征对应于模型中的一个限制条件，根据这些限制条件，选择一个满足最大熵条件的模型作为最后模型。这个原则表面上看起来和我们过去使用的原则相悖，因为本书很大篇幅都在讨论如何最小化模型对应数据的(交叉)熵的问题。但实际上，如果我们选择一个熵值相对较小的模型，其结果相当于人为地在建立的模型上增加一些信息限制条件，而导致这些限制条件的实验证据可能对我们来说是不可知的。最大熵的吸引人之处就在于熵值越大其不确定性越高。

在最大熵建模过程中，特征选择和训练通常是结合在一起的。在理想的情况下，这种方法允许我们详细地列出所有潜在的与问题解决相关的信息，而把如何得到合适模型的各种参数的任务交给训练过程。

最大熵模型存在着很多变种，本节重点介绍基于对数线性模型(loglinear model )的最大熵方法。

 (式16.4)

其中参数K表示参数的数目，为特征的权重，Z是归一化因数(通常称为“划分函数”)，确保符合概率分布的形式。为了将此模型用于文本分类，需要计算和，很简单，哪个概率值较大，就选取其为相应的类别标记。

值得注意的是，特征既包含了对象的类别信息，同时也包含了分类所需的度量信息。我们所介绍的是一种最广一泛采用的特征定义方法，即忽略对象的所属类别，只引人对象的一些基本特征信息。

式(16.4)实际上定义了一个线性对数模型。如果我们对公式的两边做对数操作，那么log p就是几个对数函数的线性组合。



在实际应用中，对数线性模型是解决分类问题的一种很重要也很常用的模型方法，其他一些常用的方法包括逻辑回归方法(McCullagh and Nelder 1989 )、分解模型( decomposable model ) ( Bruceand Wiebe 1999)以及前面章节中提到的HMM和PCFG模型。之所以引入最大熵模型，是因为近年来它已经在统计自然语言处理中获得了广泛的应用，而且它也是最大熵理论的一个最重要的应用。