

单元 8 模型评估

欢迎回到物种分布建模的在线开放课程。现在我们看看可以用来预测物种分布的不同模型，重要的是学习如何解释模型结果。建模至关重要的一步是评估模型预测结果的准确性，通常称为“验证”或“评估”。在这个单元中，我会讲如何做。

你的物种分布模型会得到多个不同的结果，你可以通过评估这些结果来确定你模型的有效性。虽然模型评估的重点通常是通过各种评估指标来衡量模型的预测性能，重要的是还要做检验可视化结果的真实情况，如预测分布图和环境变量的响应曲线。

我们先来看物种的预测分布图。我们建议你在设计物种分布模型之前，对你的物种进行一些研究，这样你能了解物种的可能分布情况。当你看物种分布模型得到的分布图，你应该严格地评估预测分布是否合理，同时考虑到扩散屏障等因素。例如，你看灰头狐蝠的预测分布图，我们预计这个物种估计会分布在西澳的一些地区，而在那里却从未观察到该物种。所以，尽管环境条件可能合适物种分布，但是可能有天然屏障阻止物种扩散到该区域。

接下来，你可以查看每个环境变量的响应曲线。这些表示在考虑输入模型其他变量的情况下，每个环境变量对应的物种分布概率。根据你对物种已知的相关知识，你可以检查这些响应曲线是否符合对环境条件的已知耐受性。例如，如果你的物种不能在高于某个温度的环境中生存，预计这一点将反映在响应曲线中。模型预测性能可以通过一套定量参数来评估，这里是指评估指标。要获得这些统计指标，首先用一组训练数据来训练模型，然后用一组测试数据来验证模型。这个验证最好用一组独立于训练数据的测试数据完成。这意味着应该不能用拟合模型的数据。一种常见方法是交叉验证。

在整个数据集的交叉验证中，这意味所有的物种分布有/无记录，被分为预定义数量的子集，也称为等份。大部分数据用来计算模型，除了一份未用于计算模型的样本数据来做模型验证。该过程的重复次数和样本份数一致，因此在本例中为 10 次，并且作为验证数据的每份子集只能使用一次。基于验证数据进行 10 次计算后的均值作为预测精度评估结果。该例子就是常用的中的 10 折交叉验证法，并且通常是软件包中运行物种分布模型的默认选择。但在理论上你可以使用任何数量的份数，如 3 等份或 5 等份。极端版本是留一交叉验证方法，重复的次数和点的数量一致，每次运行只去除 1 个数据点，剩下的点用来训练模型，并且去除的点用于验证模型。因此，交叉验证方法将各组的测试数据的运行结果进行平均，得到物种在分布图上每个位点的分布概率。大多数算法将该概率表示为范围从 0（表示低分布概率）到 1（表示高分布概率）连续值。为了计算评估指标，这些概率预测通常被转换为分类预测，即预测物种在任何给定的点上是否有分布。这个转换是基于概率预测的阈值。按照惯例，此阈值通常设置为 0.5，这意味着对于每个位置而言，概率在 0.5 以上，预测为正，表示物种有分布，某个位置的概率低于 0.5，则预测为负数，表示物种分布无。但是，有很多不同的方法来选择阈值，这取决于诸如模型的总体误差分布和分布有/无的点比率等因素。

物种分布模型分类预测结果，即物种在一个特定的地点是否有分布，可能是正确的或不正确的。将预测结果和实际的观测结果进行比较，正确的预测结果是对于真

实分布而言是真正数，对于无分布而言是真负数。有两种类型的错误：‘假正’—模型预测一个物种有分布在没有观测到的地点，‘假负’—模型预测一个物种没有分布在于其观测到的地点。模型性能可以用列联表来总结。我将解释如何根据列联表的元素中计算评估统计指标。

一个模型预测性能的简单度量是“准确度”，通过真正类和真负类相加并除以总数来简单衡量预测正确性的比例。与准确度的相反的指标是错分率，它是所有的假正类和假负类的相加再除以总数得到的。虽然这些方法很容易理解和解释，但它们并不能区别两种错误类型：假正类和假负类。另外，他们没有考虑到分布记录相对于分布无记录的比例。用一个极端的例子来说明，如果你要对有较少分布观测的稀有物种进行物种分布建模，通过预测得到所有的地点都没有分布该物种，该模型可以具有 0.9 的精度。对应的错分率为 0.1，因此只有 10% 的预测是不正确的，但存在概率是 0 的预测显然是不正确的。因此，使用更多的评估方法是真正率和假正率。

真正率是指模型正确预测物种观测出现值的比例。这是通过真正类除以真正类和假负类之和得到的值。真正率通常被称为灵敏度。与真正率相反的是假负率。高的真正率表示模型结果好。就像在这个例子中，真正率是 0.9，这就意味着假负率为 0.1。表示观测分布数据的 90% 被正确预测为物种有分布。需要注意的是，如果你只有数据分布，这两个统计数据是可以计算的。

真负率是观测的分布无样本中，模型预测正确的比例。这是通过真负类除以假正类和真负类之和得到的。真负率也被称为特异度。与真负率的相反的是假正率。同样，一个高的真负率表示模型预测性能好。如果我们再看一下这个例子，我们看到真负率为 0.96，假正率为 0.04，表明观测分布无数据的 96% 都被正确地预测。你可以通过验证真正率或者真负率是否高于期望值，检查这个预测是否具有统计学显著性。

因为列联表的值取决于概率阈值，而阈值调节评估指标的结果。阈值增加，预测分布的数据量减少。当然，观察到的分布数量保持不变，因此分布数据的正确预测比率会下降。我用 x 轴表示阈值，y 轴表示真正率和真负率的图来说明一下这个问题。当阈值增加，模型会预测出更多的分布无数据和更少的分布有数据。那么正确预测的观测分布数据的比例会下降。这意味着真正率随着阈值增大而下降，而真负率则增大。

问题是如何选择阈值来评估你的模型？有多种方法可用来选择模型阈值，我重点讲其中几个。一个常见的默认方法是设置值为 0.5，但这通常是不合适的。这种方法适合当分布有数据和分布无数据数量一样的情况，但不总是这样，因为模型正确预测分布有和分布无同等重要。在这个例子中，阈值为 0.5，真正率只有 0.12，这意味着只正确预测了 12% 的观测分布数据。另一种方法是选择一种固定的真正率，例如 95%，对应着低阈值。但此选项就会导致较低的真负率，则假正率会高。因此，更多使用的是其他可以平衡预测正确和错误的方法。例如，选择使真正率和真负率相等的阈值。这是图中两条线交叉的点。你也可以选择真正率和真负率总和达到最大时的阈值。

还有一种常用的不依赖于阈值概率的模型性能指标是相对工作特征或 ROC 图。

ROC 图是一个 x 轴是假正率，y 轴是真正率，横跨可能阈值范围的散点图。一个完美的模型只包含真正类，没有假负类。由该图中的点显示，表示假正率为 0，而真

正率为 1。包含所有可能阈值的曲线看起来像这样。模型的随机预测将出现沿着对角线从左下到右上的点。这是 ROC 空间的分隔符。任何一个在线上面的点表示预测优于随机预测，而低于线的点表示差的预测。ROC 的值是曲线下面积（AUC），是通过对 ROC 曲线下的面积累加得到的。因此，0.5 的值表示随机预测，高于 0.5 的值表示比随机预测更好的预测。ROC 曲线越接近 y 轴，曲线下的面积越大，模型越准确。一般来说，AUC 值为 0.5-0.7 被认为较低，代表模型性能差。0.7 和 0.9 之间的值被认为是合适的，高于 0.9 是优秀的模型性能。

像物种分布模型的各个方面一样，选择用来评估模型的评估指标取决于各种因素。首先，现有的评估指标取决于你在模型中使用的数据类型。有些统计数据的计算需要分布有和分布无数据，因此如果你只用分布有数据的话，则不能使用。另一件要记得的事是研究的科学问题，或者说你要如何应用模型结果。一些统计数据更好地评估物种的实际分布预测，如果你对物种的潜在分布有兴趣，例如当你研究入侵物种时，则另一些统计指标则可能更有用。正如我在较早的培训课程中已经提到的，你可能还需要对比多种算法的模型结果和评估指标，以更好地了解针对你的数据，哪种算法执行得更好。

我们已经到了关于模型评估单元的结尾。在下一个单元中，我们主要来看物种分布模式的一个应用：如何用它们与气候变化预测结合来预测未来物种分布。到时见。