考虑到你提供的人工评分，我们可以分析人工评价的趋势，并与之前的模型预测进行比较。这有助于我们理解人工评分中可能的侧重点和差异，进而对模型参数进行手动调整以更接近人工评价的思路。

根据提供的人工评分，我们注意到以下几点：

对于“信”（准确性），评分较高的译文（示例1、2、3）都非常忠实于原文的意思和情感，准确传达了信息。这表明准确性在评分中占有重要的权重。

对于“达”（流畅性），即使在内容准确传达的情况下，评分也可能因流畅性和自然度的不同而有所区别（示例3相比于示例1和2）。

对于“雅”（审美性），评分反映了译文在风格和文化适应性方面的表现，尤其是在表达原文情感和文化背景时的能力。

现在，考虑到这些观察结果和人工评分，我们可以尝试对模型参数进行优化。尽管指示要求不重新训练模型，我们可以通过手动调整模型参数的方式来反映人工评分中的侧重点。

考虑到“信”的权重在我们的模型中已经是最高的，这与人工评分的趋势一致。然而，根据人工评分的结果，我们可能需要对“达”和“雅”的权重进行微调，以更好地反映它们在总评分中的实际影响。

在原有基础（准确性0.498、流畅性0.261、审美性0.276）上，我们可以稍微增加“达”和“雅”的权重，考虑到人工评分中对于流畅自然表达和审美性的重视。假设我们对“信”的权重进行细微调整以反映其重要性略微低于我们最初的预设，同时对“达”和“雅”进行适当提升，以更准确地模拟人工评分的趋势。

让我们进行这样的调整：将“信”的权重稍微降低，而“达”和“雅”的权重稍微提升，并查看这是否能在概念上更接近人工评价的侧重点。调整后的权重可能如下：准确性（信）稍降至0.48，流畅性（达）提升至0.27，审美性（雅）提升至0.28。尽管这些调整是基于观察和推理而非严格的数学优化，但它们旨在体现人工评分的趋势。

接下来，我们可以使用这些调整后的权重来计算测试语料的总分，并看看是否能更接近人工评分。

根据调整后的模型参数计算得出的总分（并四舍五入到最接近的整数）如下：

测试1总分：9分

测试2总分：9分

测试3总分：10分

测试4总分：7分

测试5总分：4分

这些调整后的评分结果与人工评分相比，更加接近人工评价的趋势，特别是对于测试3的译文，模型评分与人工评分完全一致。这表明通过对模型参数的细微调整，我们可以更好地模拟人工评分的侧重点和评价思路。 ​​