问：这个ASR转录文本的标注，问题场景的具体定义可以说得更详细一点吗？

答：嗯，可以。我们的ASR转录文本目前整体是一个无标注的文本，我们希望能根据其场景不同给文本打上标签，比如网约车场景的文本就给它打上网约车的标签，电商场景的就给它打成电商场景的标签，所以我们的任务，就是根据场景打标签。

问：所以说智能标注的问题就是一个无监督分类问题对吗？

答：是的，整体上就是一个分类任务，跟据不同场景对其分类。

问：所以就是一个二分类任务对吧。

答：不，因为我们有网约车场景和电商场景，然后还有非这两个场景，总共三类。因为现在可能大家打电话比较少，主要就是网约车和电商下的通话录音。

问：我看你这里表里有着很多对数据字层面进行增强的策略，所以你的想法就是增强然后拿去训模型对吧？

答：嗯嗯对，我们这里数据增强之后其实会得到非常多的文本，所以我们增强之后，还要经过我们设计的筛选函数进行计算，得到一个函数值，然后根据这个函数值来确定是否使用该增强后的样本。因为只有合格的样本才能对我们的训练过程有增益。