



自动驾驶汽车 预测-决策-规划-控制实战入门

2.2 构建S型轨迹预测的数据集

创作者: Ally

时间: 2021/11/14





1 训练集、测试集与验证集的区别

- ◆ 训练集——课本：学生根据课本里的内容来掌握知识。
- ◆ 验证集——作业：通过作业可以知道不同学生学习情况优劣、进步的速度快慢。
- ◆ 测试集——考试：考的题是平常都没有见过，考察学生举一反三的能力。

类别	训练集	验证集	测试集
是否被训练	是	否	否
作用	训练模型参数（如神经网络连接权重）	调超参数（如神经网络隐层数），监控模型是否发生过拟合（以决定是否停止训练）	为了评估最终模型泛化能力
使用次数	多次使用，以不断调参		仅仅一次使用
样本占比	训练集+验证集的90%左右	训练集+验证集的10%左右	视需泛化测试的数据样本大小，与训练集和验证集数量无关
缺陷	模型在一次次重新手动调参并继续训练后所逼近的验证集，可能只代表一部分非训练集，导致最终训练的模型泛化性能不够		测试集为了具有泛化代表性，往往数据量比较大，测试一轮要很久，所以往往只取测试集的其中一小部分作为训练过程中的验证集

◆ 利用Matlab自动驾驶仿真获取数据集主要思想：

- 利用Matlab自动驾驶工具箱的road函数，构造不同曲率半径的S型道路；
- 然后利用roadBoundaries函数获取该道路的左右两条边界，并对边界坐标索引号重新整理；
- 根据单向两车道约定及边界坐标，计算左车道中心线坐标，以生成2号交通车行驶路径点
- 考虑到普通的路径点无时间信息，无法满足轨迹预测要求，再利用该路径点和交通车速度基于trajectory函数为交通车赋予轨迹信息；
- 目前暂未研究清楚从何处可以直接调取生成的轨迹，因此考虑利用advance函数进行仿真，获取交通车每一个时间步的轨迹坐标序列。

◆ 后续处理流程

- 随着S型道路不同，获得的XY轨迹坐标差异较大，这种差异会直接体现在LSTM网络训练的精度上，故还需要进行标准化处理；
- 利用mean函数和std函数获取数据集的平均值和标准差，进而转化为具有零均值和单位方差的标准数据集；
- 同时，后期在LSTM网络预测时，为方便逐步训练，利用cell类型数据类型存放上述若干数据集。

