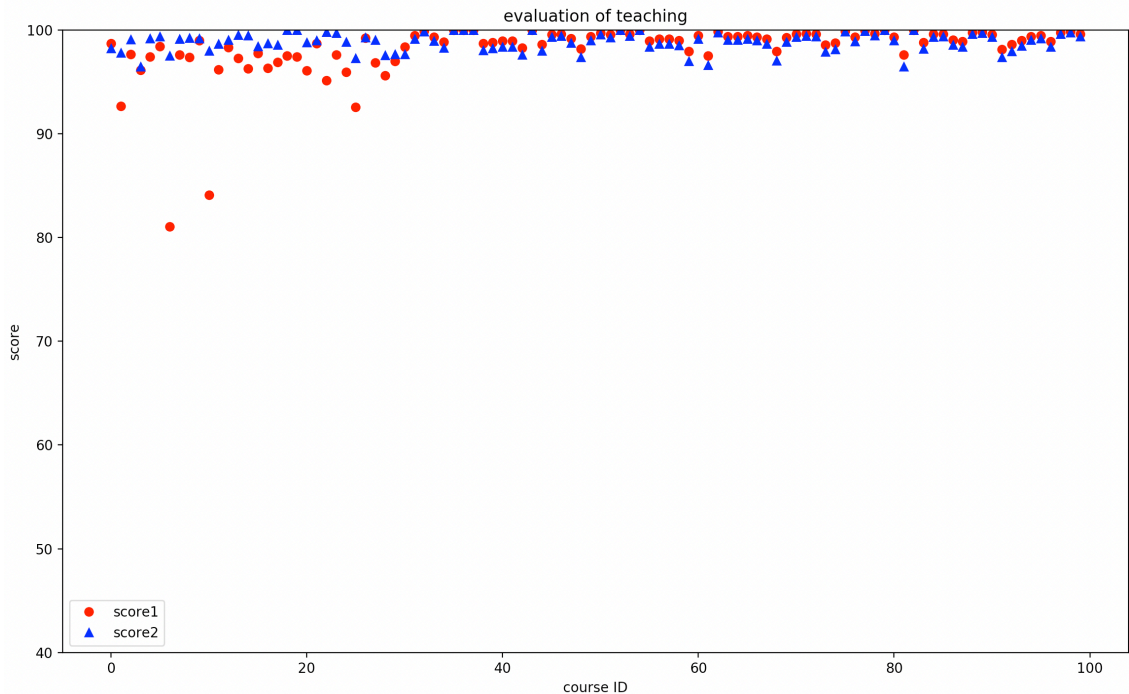


上周工作

- 整理数据：201801学期的教师测评中，全部学生参与教评的数据有1642条，存在未参加教评的数据有1628条；选取100条数据（前30条不完备的数据和后70条完备的数据），得到用ER算法得到的教评结果score1和采用平均值计算得到的教评结果score2的比较，



可以看到，后面数据两种方法相差不是很多，前面数据中，score1和score2存在较大差别，例如，从左到右第七个点，相差特别大，是因为上课人数远多于教评人数（上课26人，教评8人），教评结果中存在较大不确定性，拉低了教师的得分。

- 阅读张捷学姐的毕业论文，最后一篇的论文中的那个两阶段离散化方法的使用看不大懂。感觉DBRB对与缺失部分数据的数据集具有更强的适应性，不知道能否找个方法，改进DBRB的规则库构建步骤，使DBRB通过不完整的数据集构建完规则后，仍有较强的推理能力

下周工作

- 解决教评数据的预测模型，当前问题主要有：1.没有合理的指标能够最为用来预测的属性；2.没法构建规则库，之前想的是通过第一步的ER方法，能够得到教师的一个合理的分布式评价结果，但没有考虑前提属性应该如何选取；3.未能充分的利用数据，手头中的教评数据只用在了ER方法中，之前的想法是用教评的历时数据进行预测，但可能会面临某门新开设的课程无法预测的情况，另一个想法就是预测还是使用当前的教评中的指标，然后考虑构建一个DBRB系统，例如张捷学姐论文中的使用聚类算法去构建规则库。
- 理解两阶段离散化方法在BRB中的应用