

- ▶ **经皮冠状动脉介入治疗 (percutaneous coronary intervention, PCI)**是指经心导管技术疏通狭窄甚至闭塞的冠状动脉管腔，从而改善心肌的血流灌注的治疗方法。
- ▶ **无复流现象 (no-reflow phenomenon)**是指冠状动脉闭塞，血流中断后重新恢复血流，却无心肌组织的有效灌注的现象。
- ▶ **研究复流现象的意义**: 通过病人的指标来判断病人手术是否足够安全，通过这种判断方式启发性的改进我们的术前方案，针对不同类型的人群采取不同的治疗策略以及术前身体调整。

数据预处理——缺失数据填补

一、变量说明

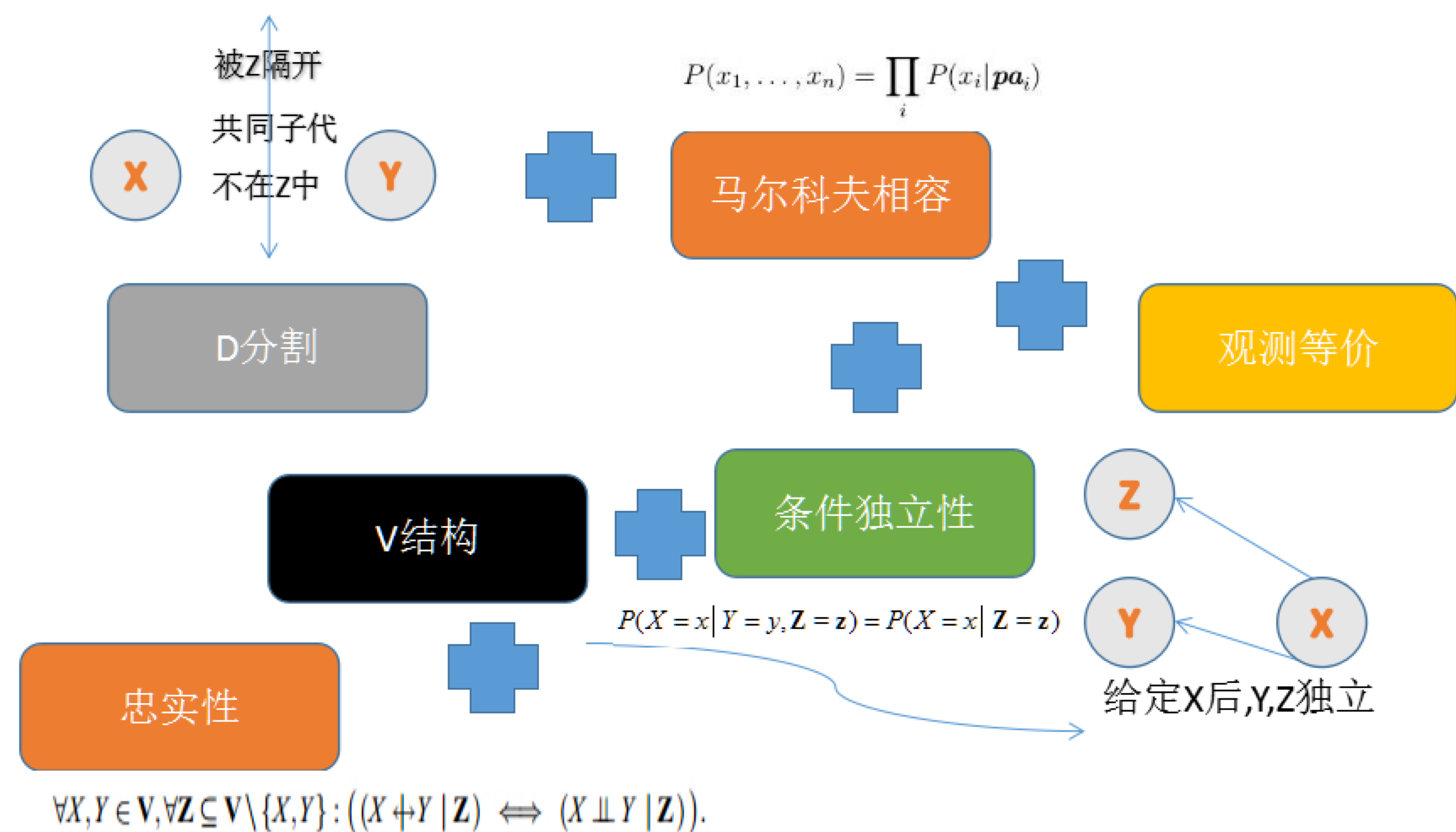
- PCI 数据集中可以分为术前指标，术中指标，和术后指标。其中术前指标主要指的是基本信息，病史，用药史等。术中和术后指标指的是在 PCI 手术开始之后的一系列测量指标。

二、算法简介

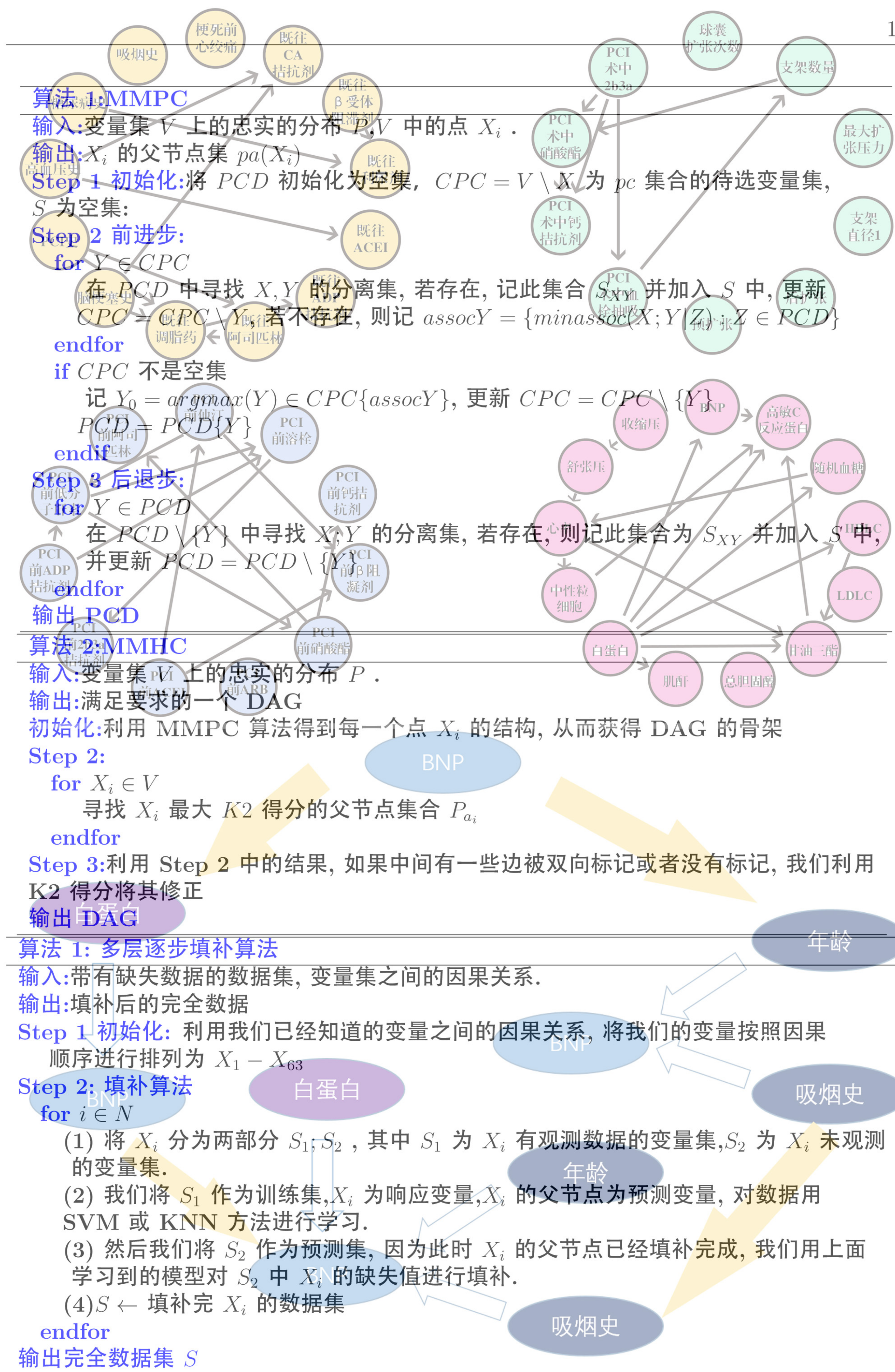
- 我们主要通过 *The Max-Min Hill-Climbing Bayesian Network Structure Learning Algorithm* (Tsamardinos 2006)（以下简称 MMHC）算法来分别对我们分层数据进行学习。
- MMHC 属于 DAG 混合算法，具有精确稳定的特点。它先通过 MMPC 算法来学习 DAG 的估价，然后用贪心法和 K2 得分来在结构约束的空间下选择最合适的 DAG。

三、有向无环图 (DAG) 简介

- 我们用有向无环图用来表示变量之间的因果关系，例如 $X \rightarrow Y$ 来表示变量 X 可以影响变量 Y ，满足相容性的分布 P 和有向无环图 G 结合起来构成了一个贝叶斯网络 $\langle G, P \rangle$ 。关于贝叶斯网络。



数据缺失填补算法

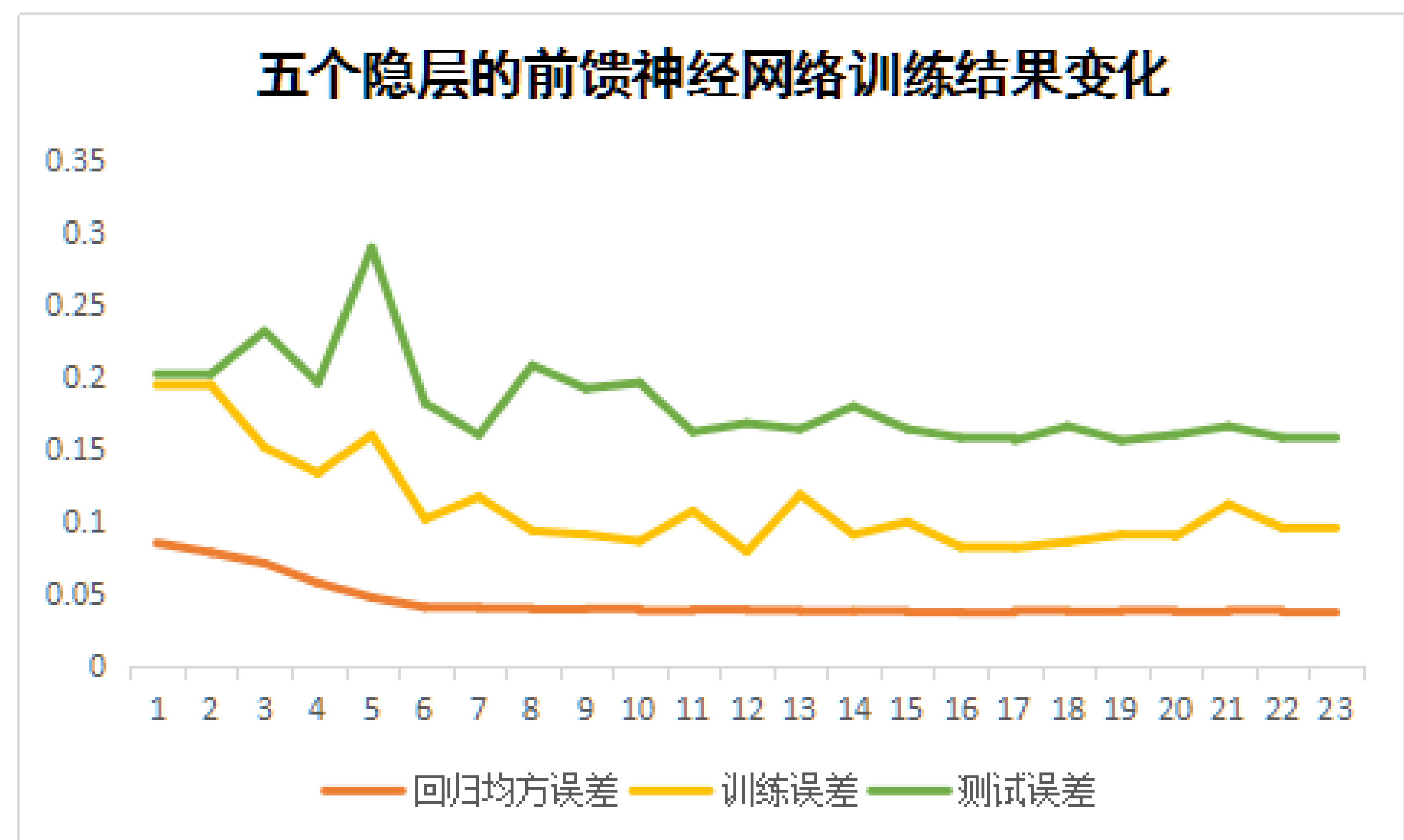


说明

苗旺和李艳芳 (2013) 按照时间顺序分为六层建立了 PCI 数据缺失机制模型, 并用 Logistic 回归取得了良好的效果。但是在他们的研究中并没有考虑到每一层数据之间的相互作用, 在不考虑数据的层间作用的前提下, 他们的估计可能不够精准, 他们的缺失模型可能并不完善。

我们分别针对全部指标和仅术前指标, 对是否会出现术后复流现象, 进行了预测预测。其中, 因为原始数据缺失较多, 我们通过 DAG 模型, 进行了多层逐步填补。并和随机填补的结果进行比较。右边是我们学习的模型的结果。

算法结果展示



术前指标训练结果:

	Boosting	随机森林	Bagging
随机填补	84.80%	87.27%	87.27%
多层填补	86.24%	82.96%	83.57%

全部指标训练结果:

	Boosting	随机森林	Bagging
随机填补	88.71%	90.55%	87.06%
多层填补	91.17%	90.76%	91.58%

结论

数值结果表明, 多层逐步填补在全部指标的预测上比随机填补略有提高。通过机器学习的方法, 我们可以较好的预测一个患者在术前是否会出现复流现象, 我们从而可以通过分析患者的物理条件来判断他是否适合进行手术。通过因果结构的学习, 我们也了解患者应该改善哪些身体条件, 进行有效可靠的术前治疗。

参考文献

- (1) 自变量有缺失的分类问题 李艳芳, 苗旺
- (2) I. Tsamardinos, L.E. Brown, C.F. Aliferis. The max-min hill-climbing Bayesian network structure learning algorithm[J]. Machine learning. 2006, 65(1):31-78
- (3) 通过术前指标对经皮冠状动脉介入治疗术后无复流现象的预测 吕渊