# 统计关系学习

独立同分布iid、统计关系学习、迁移学习这三者是互不相同的。

传统机器学习模型假设数据是独立同分布的 (independent and identically distributed, iid)，也就是说，数据样本之间相互独立，不存在任何关系。然而，实际上有许多都是有关系的，比如学术论文之间存在引用关系等。我们把这种样本之间存在关系的数据叫做关系数据 (relational data)，把基于关系数据的机器学习叫做统计关系学习 (statistical relational learning, SRL)。

## 1. 主流的统计关系学习方法

主流的统计关系学习方法可以大致分为以下五类：

1）基于个体推理模型的方法(IIM)

从关系信息中抽取出特征，并将关系数据转化成适合传统机器学习算法的特征向量形式，然后用传统学习算法来进行分类。

由于IIM方法忽略了关系数据中样本之间的相关性，分类性能往往比较差。另外，这类模型主要用于分类任务，而不适合于链接预测等其它统计关系学习任务。实用性不强

2）基于启发式联合推理的方法（HCI）

采用启发式的循环迭代方式对多个样本同时进行分类。由于考虑了有关系的样本之间的相关性，这类方法在实际应用中取得了比IIM方法更好的分类性能。 但是，HCI方法主要用于联合分类，而不适合于其它统计关系学习任务。因此，HCI方法的实用性也比较有限。

3）基于概率关系模型的方法（PRMs）

通过拓展传统的图模型来对样本之间的相关性进行建模。

典型的PRMs包括关系型贝叶斯网络（RBNs）、关系型马尔可夫网络(RMNs)和关系型依赖网络(RDNs)。

结构学习是对变量之间的各种依赖关系进行确定，是一个组合优化问题，因此复杂度非常高。另外，RMNs和RDNs的参数学习也没有收敛和快速的方法，实际应用中往往采用一些近似的逼近策略。因此，PRMs方法的一个主要缺点是学习速度慢，只适合于小规模数据的处理。

4）基于概率逻辑模型的方法（PLMs）

将概率引进一阶谓词逻辑，能够很好地对关系数据进行建模。

代表性的PLMs有概率Horn溯因(PHA),贝叶斯逻辑编程(BLP),马尔可夫逻辑网络（MLNs）。

PLMs的底层建模工具还是基于图模型，因此基于概率逻辑模型的方法（PLMs方法）也具有PRMs方法的学习速度慢的缺点，只适合于小规模数据的处理。

5）基于隐因子模型的方法(LFMs)

将统计学中的隐因子模型引进统计关系学习。其中，常用的矩阵分解算法对应于某个LFM的一个最大似然估计或者最大后验估计。因此，矩阵分解方法可以看成LFMs的一种特例。

大部分LFMs方法具有相对于观察到的链接数的线性复杂度，学习速度远远超过PRMs方法和PLMs方法，能很好地对较大规模数据进行建模。因此，包括矩阵分解这个特例在内的LFMs方法已经发展成为目前统计关系学习算法的主流，具有比其他方法更广阔的应用前景。

## 2. 趋势和热点

1）将统计关系学习理论和方法应用到知识图谱(knowledge graph)是最近发展起来的一个新的研究热点。

2）设计在线学习模型以实现对动态关系数据的有效建模是统计关系学习的一个研究热点。

3）对这些超大规模数据（大数据或海量数据）的分析和处理将给统计关系学习研究带来极大的挑战。因此，设计超大规模分布式学习算法以实现对海量关系数据的有效建模是另一个值得深入探索的研究热点。

## 3. 统计关系学习的应用

    （1）命名实体的识别，判断相同名字的是不是同一个人

    （2）金融欺诈：想法就是和诈骗犯相关的人，估计也是诈骗犯

    （3）在dblp数据上，预测怎么洋的文章容易被发表和引用

    （4）判断电影在上映后的一个星期里，会不会票房达到100万（比如某明星，一部电影走红，下一部电影走红的可能性相当大。）

## 参考资料

https://blog.csdn.net/ictextr9/article/details/4724010