基于集成学习的社区发现算法研究

组长：黎家昊

组员：黄晋涛　史晨阳　邱凯

报告内容

① • 课题研究的背景及意义

② • 课题研究目标、研究内容及关键内容

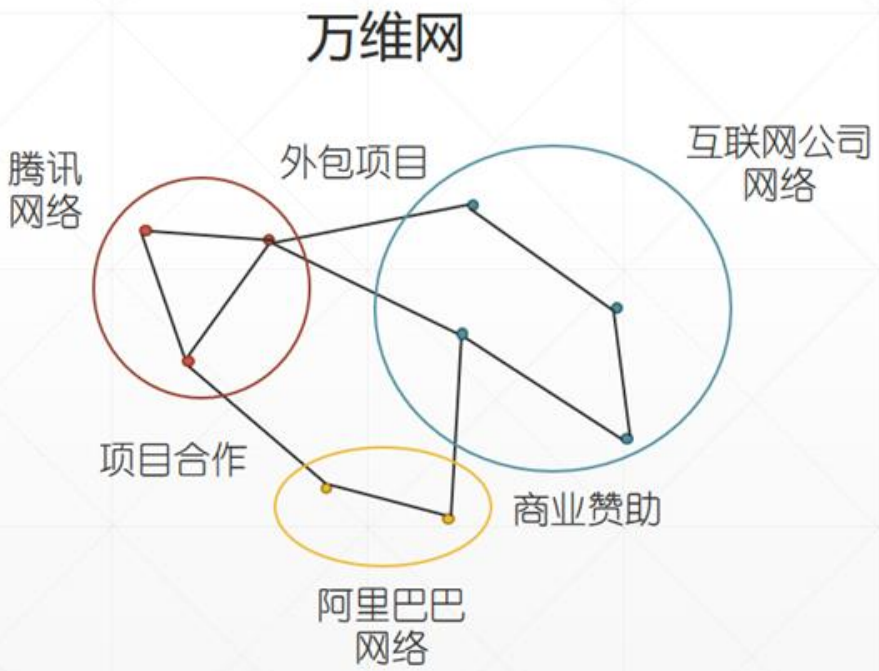
③ • 拟采取的技术路线

④ • 课题的创新性

⑤ • 计划进度和预期成果

① 研究背景与研究意义

我们所在的现实世界中，宏观到微观都可以看成一个由若干相互作用的子系统组成的复杂系统，把子系统抽象为节点，把子系统间的相互作用抽象成连接节点就构成了复杂网络。



社区发现研究社会网络中社区结构和功能，对于认识网络功能等有深刻的意义和实际应用价值。

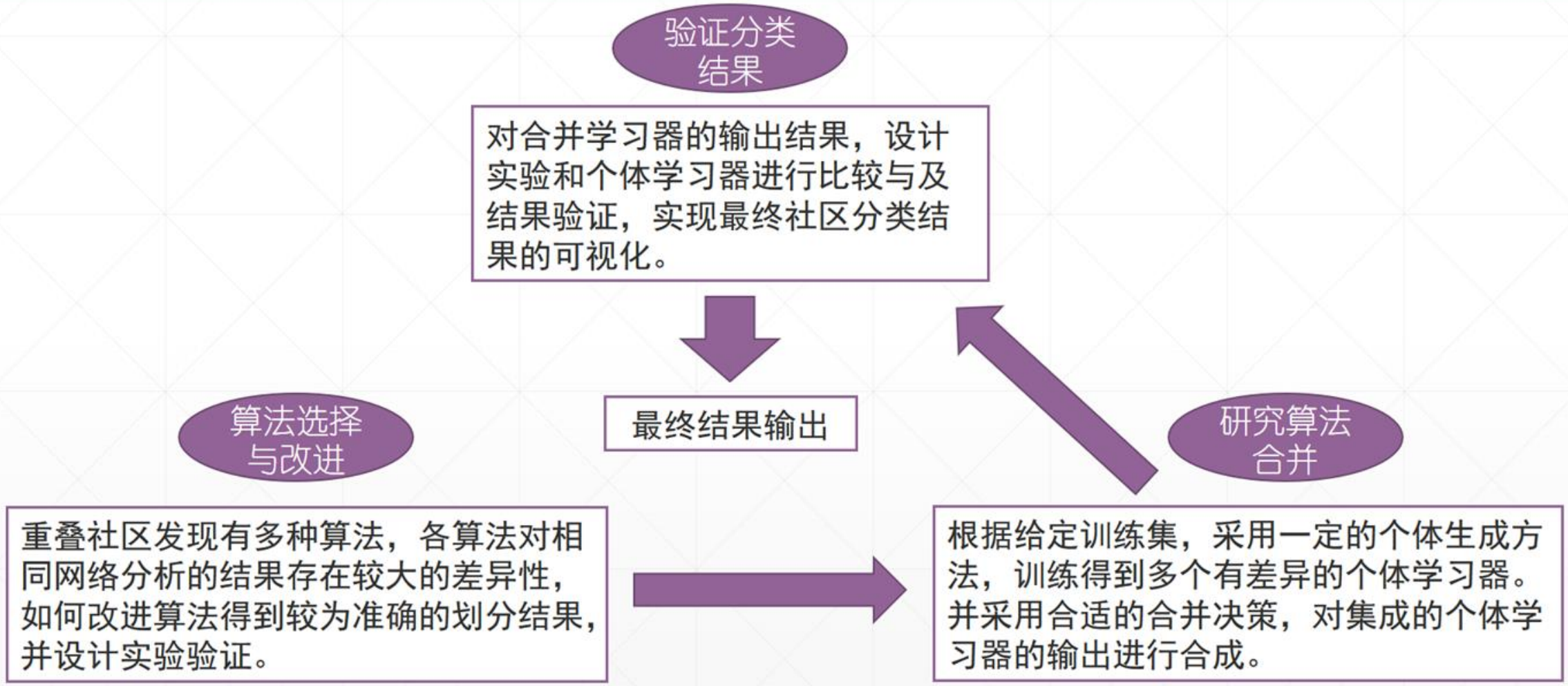
② 研究目标

①
- 多类社区发现算法的实现

②
- 利用集成学习融合多种算法分析网络

③
- 完成实验以及可视化系统

② 研究内容


验证分类结果
对合并学习器的输出结果，设计实验和个体学习器进行比较与及结果验证，实现最终社区分类结果的可视化。
最终结果输出
算法选择与改进
重叠社区发现有多种算法，各算法对相同网络分析的结果存在较大的差异性，如何改进算法得到较为准确的划分结果，并设计实验验证。
研究算法合并
根据给定训练集，采用一定的个体生成方法，训练得到多个有差异的个体学习器。并采用合适的合并决策，对集成的个体学习器的输出进行合成。

② 关键问题

针对多种类型的原始数据，如何选择相应的社区发现算法的问题

复杂网络中非数值的信息转化为数值信息，评估缺少语义信息的网络

设计实验过程中数据集归一化以及其他问题

如何选择合适的合并决策，对各类社区发现算法进行集成

③ 拟采取的技术路线

以微博为例，微博下的可用信息有：

- 地理位置
- 毕业院校
- 标签信息
- 微博内容
- 关注
- 粉丝
- ……

③ 拟采取的技术路线

主要研究的算法：

对彼此交叉、相互重叠的网络社区进行分析时。基于标签传播、基于团渗理论、基于链接划分、基于局部扩展、基于代理和动态方法、基于模糊检测等重叠社区结构的发现算法有比较多的关注和研究。

SLPA算法

Algorithm	Year	Complexity
CFinder	2005	
LPA	2007	$O(m)$
LFM	2009	$O(n^2)$
EAGLE	2009	$O(n^2 s)$
GIS	2009	$O(n^2)$
HANP	2009	$O(m)$
GCE	2010	$O(mh)$
COPRA	2010	$O(vm \log(\frac{vm}{n}))$
NMF	2010	$O(Kn^2)$
Link	2010	$O(nk_{max}^2)$
SLPA	2011	$O(Tm))$
BMLPA	2012	$O(n \log n))$

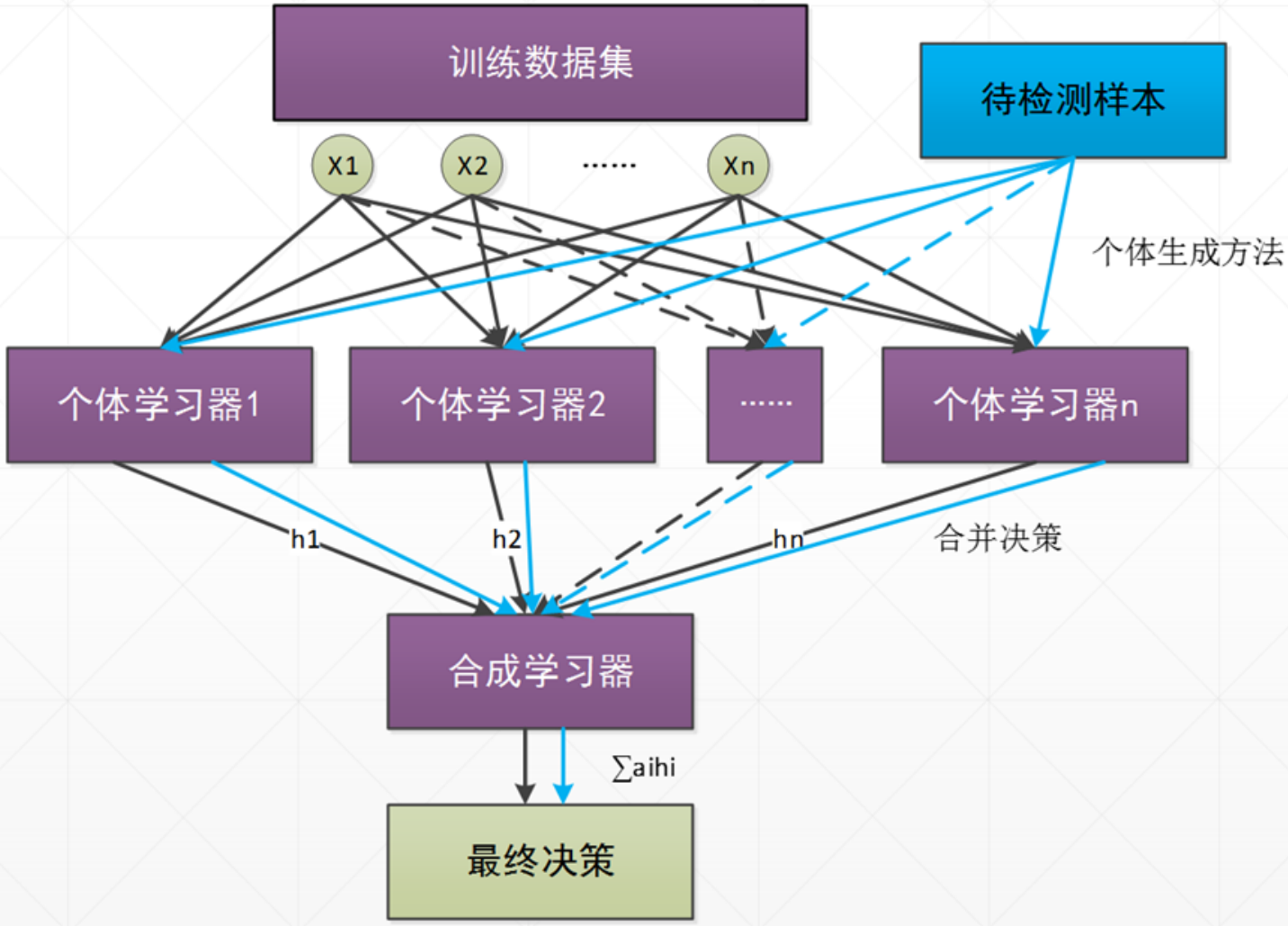
③ 拟采取的技术路线

集成学习融合算法：

多年来的研究表明，通过集成学习得到的学习模型要明显优于个体学习模型，且能有效提高模型的泛化能力。

集成也分几种：

- 不同算法的集成
- 同一种算法在不同设置下的集成
- 数据集的不同部分分配给不同分类器之后的集成

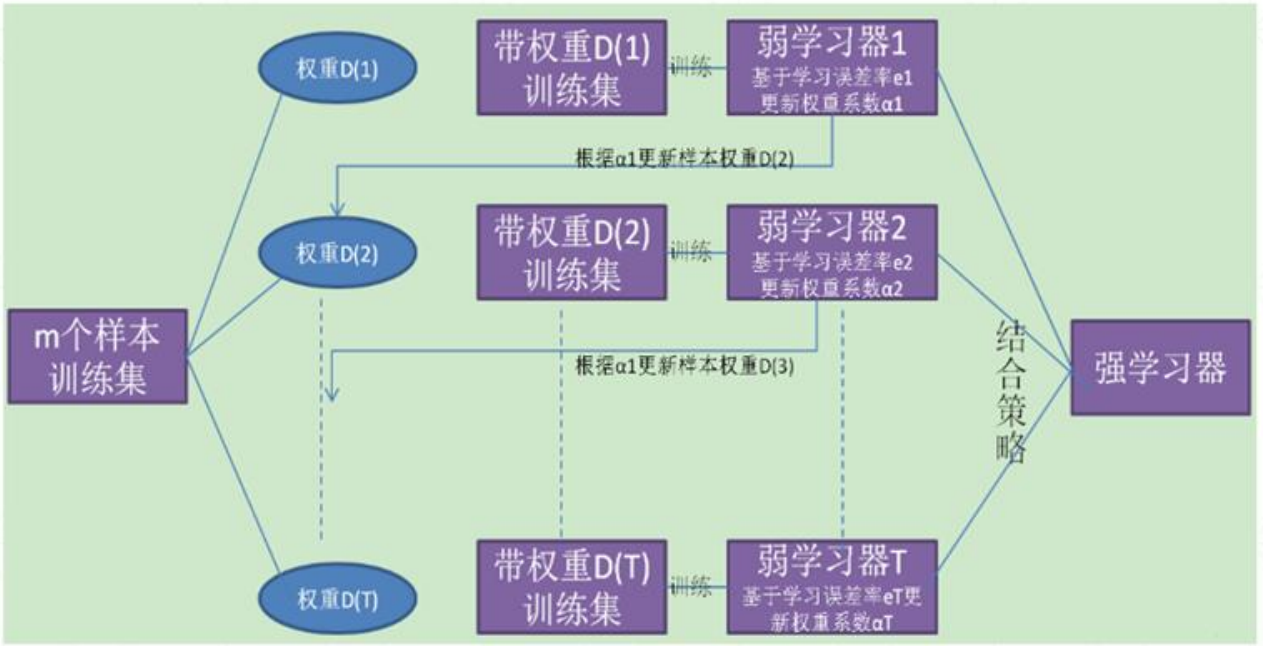


③ 拟采取的技术路线

个体学习器生成：

代表的方法有：boosting、bagging、随机森林算法。

例如右图针对个体学习器之间存在强依赖关系，一系列个体学习器基本都需要串行生成的boosting系列算法



合并方案：

对于数值类的回归预测问题，通常使用的结合策略是平均法。

对于分类问题的预测，通常使用的是投票法。

学习法，如stacking，不是对弱学习器的结果做简单的逻辑处理，而是再加上一层学习器，将训练集弱学习器的学习结果作为输入，将训练集的输出作为输出，重新训练一个学习器来得到最终结果。

④ 课题的创新性

- 应用多种算法，并进行算法之间效率以及准确度的比较。
- 优化并改进传统的网络社区发现算法。
- 编程在实际网络社区中爬取数据，数据真实可靠。
- 利用集成学习综合多种算法分析复杂网络重叠社区，使分析更加准确。

⑤ 计划进度和预期成果

起止时间	开展内容	预期成果
2017年4月 – 2017年6月	项目初步实施，小组进行理论知识的学习研究	了解各算法的思想以及实现步骤，初步编程实现算法
2017年6月 – 2017年9月	研究论文，设计重叠社区的发现算法	实现重叠社区发现算法并对算法进行改进
2017年9月 – 2017年12月	研究并利用集成学习综合多种算法分析网络	实现各算法的个体学习器并选择合适的合并策略
2018年12月 – 2018年2月	编程实现社交网络上数据的爬取	实现能够爬取实际网络数据并进行分析的功能
2018年2月 – 2018年3月	分析结果的可视化与软件设计	完成可视化软件的封装
2018年3月 – 2018年4月	完成研究论文，准备参加结题答辩	撰写论文、结题报告与答辩PPT

Thank You