**选题报告**

计14 2011011279 汪忆辰

计15 2011011300 周若凡

计15 2011011308 叶 紫

1. **选题内容**

通过对社交媒体（如微博、人人）中的文本评论、音频数据、图片数据、视频信息（URL、标签、评论等）为基础，分析用户的性格特点。

1. **选题背景**

现在使用社交网络的人越来越多，而一些学术研究表明人们在社交网络上的一些行为和自身的性格有一定的联系。我们希望能通过对社交媒体上各种信息的分析给予用户性格上的一个预测，进而可以获得一些统计意义上的结论。

1. **探究基础**

现有的各种paper，如：

Personality in Cyberspace: Personal Web Site as Media for Personality Expressions and Impressions

The intersection of users’ personality and social media use

等等。

1. **方案初稿**

我们构想设计一个模型，这个模型可以对用户社交网络上发表和分享的文本、图片和视频音频等进行综合性的评分，并给予这个用户大五类人格特性各项指标的一个分布情况估计分析。

我们会先在网上搜索相关论文，搜索关于文本、图片、视频音频等相关分析用户特性的模型或是算法（关键字抓取、音频性格分析、以及表情统计等等）；同时利用标准的心理问卷对周围使用社交网络的人进行五型人格的测试已经对他们社交网络的网址进行登记（经过他们同意后）用作标准集；在数据和算法的基础上寻找创新点、构建一个我们自己的模型，并初步规划用机器学习的方法调整参数让模型测试结果更符合我们收集的数据。最终我们希望能在社交网络上构建一个类似的APP供用户体验并反馈结果。

社交网络我们暂限制于人人网（www.renren.com）。

数据收集范围也暂定在本系以及周围的好友同学。

1. **进度安排**

前期准备在第8周周末完成（包括收集标准集数据、算法的寻找）

模型部分完成：基本代码在第10周周中完成；之后1周利用数据调整参数和模型。

最后的时间完成在社交网络（人人）上的相关应用制作。

1. **预期成果**

能够通过对人人网上信息的抓取给出一个较为科学的用户的性格分析（五型性格各自的分数值），并制作成一个应用给用户娱乐。