**信号处理原理课程SRT工作进度报告**

**——基于人人网多媒体数据的用户性格分析**

**2011011300 计15 周若凡**

**2011011308 计15 叶紫**

**2011011279 计14 汪忆辰**

1. **总体任务**

通过对人人主页中的文本评论、音频数据、图片数据、视频信息（URL、标签、评论）等多媒体数据的处理和分析，结合机器学习方法，按照大五类人格模型分析用户的性格特点。

1. **本阶段完成任务**
2. 借用实验室的机器，采用django的web框架搭建数据采集服务器。
3. 完成用于数据采集的人人APP，为用户提供大五类人格测试服务，同时经用户同意后采集用户主页的多媒体数据，包括状态、评论、语音和图像相册、个人资料等数据。
4. 将采集数据的人人APP提交到人人网上，在好友圈中宣传，采集到一定数量的样本。
5. **数据采集方法**

经徐明星老师同意，我们借用了实验室的机器作为前期数据采集和后期成果展示APP的服务器。我们应用了大一暑期小学期学习的django的web应用框架，编写了一个用于采集人人用户信息的APP，主要功能为：

1. 通过让用户填写一份标准的大五类人格测试量表，将填写的内容post到大五类人格测试的权威网站上（<http://www.outofservice.com/bigfive/>）。该网站提供免费、标准、权威的大五类人格测试服务，并返回五类人格特性的百分比评分结果。我们将该网站计算的结果抓取回来，展示给测试用户，并在本地保存。
2. 我们在经过用户同意后，将采集用户id下的所有多媒体数据，包括用户的状态、评论、相册（包括语音相册）、视频等，这些多媒体数据将作为后续机器学习的数据集。我们通过用户id保证人格测试的标准结果能够和该用户的个人数据一一对应。
3. **下一步任务计划**

从权威人格测试网站上获取的大五类人格测试结果可以认为是该用户性格的标准结果，我们将在采集到足够多（不少于100份）的用户数据后，通过对数据的分析，确定和用户性格相关度较高的特征。这些特征将分为不同类别，包括文本、语音、图片、个人信息等。

然后我们将通过不同的机器学习方法（M5‘ Rules和Gaussian Processes等），得到用户性格的预测模型，分析比较不同模型的性能优劣。

最后将性能最好的预测模型集成到成果展示的人人APP中，向人人用户提供基于我们的预测模型的性格分析APP，同时和标准结果进行对比展示。