**信号处理原理课程SRT第13周进度报告**

**2011011279 汪忆辰**

**2011011300 周若凡**

**2011011308 叶紫**

这周我们主要分工完成了一下几个方面的工作：

1. 叶紫同学完成了对所有用户的信息抓取工作。同时，考虑到c、java、python等语言都有支持json的工具包，所以将数据转换为json格式，方便了下一步对数据的分析。
2. 汪忆辰同学完成了对文本类信息特征的提取，包括人人用户的状态、日志、分享和个人资料，从中提取出45个量化的特征。目前正在利用weka的机器学习工具对不同机器学习方法的参数进行调整，对训练效果进行比较。附件中为代码生成的特征数据文件。
3. 周若凡同学基本完成了对音频和图片的特征提取工作，目前正在考虑如何将这部分特征数据与文本类数据合并处理，以及最终的服务器端如何从数据抓取开始调用不同语言的代码得到预测结果。

下一步我们将对数据间的相关性进行分析，得出一些统计意义的结论；同时将文本类数据和音频、图片的特征数据合并后再进行训练。最终版本的人人App也将开始着手开发。

目前用GaussianProcess模型得到的训练结果最好，一下为部分训练结果：

1、Agreeableness 宜人性：

Options: -L 1.0 -N 0 -K "weka.classifiers.functions.supportVector.RBFKernel -C 2

50007 -G 1.0"

Gaussian Processes

Kernel used:

RBF kernel: K(x,y) = e^-(1.0\* <x-y,x-y>^2)

Average Target Value : 0.6319493670886076

Inverted Covariance Matrix:

Lowest Value = -0.21224204798424406

Highest Value = 0.7131644310303128

Inverted Covariance Matrix \* Target-value Vector:

Lowest Value = -0.20683938659566412

Highest Value = 0.18733234042459573

Time taken to build model: 0.33 seconds

Time taken to test model on training data: 0.02 seconds

=== Error on training data ===

Correlation coefficient 0.9632

Mean absolute error 0.0634

Root mean squared error 0.0813

Relative absolute error 55.2695 %

Root relative squared error 56.0033 %

Total Number of Instances 79

=== Cross-validation ===

Correlation coefficient -0.0033

Mean absolute error 0.1158

Root mean squared error 0.1473

Relative absolute error 99.3901 %

Root relative squared error 99.6487 %

Total Number of Instances 79

2、Openness 开放性：

Options: -L 1.0 -N 0 -K "weka.classifiers.functions.supportVector.RBFKernel -C 2

50007 -G 1.0"

Gaussian Processes

Kernel used:

RBF kernel: K(x,y) = e^-(1.0\* <x-y,x-y>^2)

Average Target Value : 0.6203670886075948

Inverted Covariance Matrix:

Lowest Value = -0.21224204798424406

Highest Value = 0.7131644310303128

Inverted Covariance Matrix \* Target-value Vector:

Lowest Value = -0.2341211543013501

Highest Value = 0.24012971361079896

Time taken to build model: 0.31 seconds

Time taken to test model on training data: 0.03 seconds

=== Error on training data ===

Correlation coefficient 0.9701

Mean absolute error 0.0706

Root mean squared error 0.0871

Relative absolute error 60.1676 %

Root relative squared error 59.1098 %

Total Number of Instances 79

=== Cross-validation ===

Correlation coefficient -0.3837

Mean absolute error 0.1293

Root mean squared error 0.1588

Relative absolute error 107.1496 %

Root relative squared error 105.1369 %

Total Number of Instances 79

1. Neuroticism 神经性：

Options: -L 1.0 –N 0 –K “weka.classifiers.functions.supportVector.RBFKernel –C 2

50007 –G 1.0”

Gaussian Processes

Kernel used:

RBF kernel: K(x,y) = e^-(1.0\* <x-y,x-y>^2)

Average Target Value : 0.5150253164556963

Inverted Covariance Matrix:

Lowest Value = -0.21224204798424406

Highest Value = 0.7131644310303128

Inverted Covariance Matrix \* Target-value Vector:

Lowest Value = -0.2861626264978455

Highest Value = 0.2565243701384649

Time taken to build model: 0.27 seconds

Time taken to test model on training data: 0.25 seconds

=== Error on training data ===

Correlation coefficient 0.9661

Mean absolute error 0.0764

Root mean squared error 0.0972

Relative absolute error 56.4127 %

Root relative squared error 55.9919 %

Total Number of Instances 79

=== Cross-validation ===

Correlation coefficient 0.0804

Mean absolute error 0.1382

Root mean squared error 0.1737

Relative absolute error 100.4282 %

Root relative squared error 98.7137 %

Total Number of Instances 79

1. Conscientiousness 严谨性：

Options: -L 1.0 –N 0 –K “weka.classifiers.functions.supportVector.RBFKernel –C 2

50007 –G 1.0”

Gaussian Processes

Kernel used:

RBF kernel: K(x,y) = e^-(1.0\* <x-y,x-y>^2)

Average Target Value : 0.5253417721518988

Inverted Covariance Matrix:

Lowest Value = -0.21224204798424406

Highest Value = 0.7131644310303128

Inverted Covariance Matrix \* Target-value Vector:

Lowest Value = -0.16567229015318807

Highest Value = 0.16736739294886496

Time taken to build model: 0.25 seconds

Time taken to test model on training data: 0.02 seconds

=== Error on training data ===

Correlation coefficient 0.9708

Mean absolute error 0.0623

Root mean squared error 0.075

Relative absolute error 52.5162 %

Root relative squared error 51.8927 %

Total Number of Instances 79

=== Cross-validation ===

Correlation coefficient 0.2784

Mean absolute error 0.1166

Root mean squared error 0.1405

Relative absolute error 95.9551 %

Root relative squared error 95.3594 %

Total Number of Instances 79

5、Extraversion 外向性：

Options: -L 1.0 -N 0 -K "weka.classifiers.functions.supportVector.RBFKernel -C 2

50007 -G 1.0"

Gaussian Processes

Kernel used:

RBF kernel: K(x,y) = e^-(1.0\* <x-y,x-y>^2)

Average Target Value : 0.5280506329113923

Inverted Covariance Matrix:

Lowest Value = -0.21224204798424406

Highest Value = 0.7131644310303128

Inverted Covariance Matrix \* Target-value Vector:

Lowest Value = -0.16676965438423644

Highest Value = 0.16665276126472978

Time taken to build model: 0.26 seconds

Time taken to test model on training data: 0.02 seconds

=== Error on training data ===

Correlation coefficient 0.9591

Mean absolute error 0.0616

Root mean squared error 0.0774

Relative absolute error 57.4631 %

Root relative squared error 56.8559 %

Total Number of Instances 79

=== Cross-validation ===

Correlation coefficient -0.0553

Mean absolute error 0.1098

Root mean squared error 0.1392

Relative absolute error 101.2316 %

Root relative squared error 100.9279 %

Total Number of Instances 79