**深 圳 大 学 实 验 报 告**

**课程名称： 面向对象技术与方法**

**实验项目名称： 实验4数据预处理与类信息统计**

**学院：**

**专业：**

**指导教师： 尹剑飞**

**报告人： 学号： 班级：**

1. 实验要求
2. 数据预处理

本次实验在实验1和实验2的基础上，增加如下需求：

1. 对附入的数据集D中每一条记录Ri的每一个字段Vi（除去标签、日期字段）进行：，规范化操作，要求支持以下规范化方法：
   1. Z-score scaling: variables recalculated as (Vi - mean of {Vi})/s, where "s" is the standard deviation of {Vi}.As a result, all variables in the data set have equal means (0) and standard deviations (1) but different ranges.
   2. Dividing standard deviation: Dividing each value Vi by the standard deviation of {Vi}. This method produces a set of transformed variables with variances of 1, but different means and ranges.
   3. 0 - 1 scaling: for each variable Vi: (Vi - min {Vi}) / (max {Vi} – min{Vi})
   4. Dividing each value by the range: for each variable Vi: Vi / (max {Vi} - min {Vi}). In this case, the means, variances, and ranges of the variables are still different, but at least the ranges are likely to be more similar.

因为输入的每一个数据集D中每条记录的每一个字段的数值精度可能不一样，要求程序找出该数据集D的最大精度，同时以该最大精度确定上述规范化后的字段精度。

1. 对一个输入的数据集D，在执行完实验1所要求的错误检查后，若数据集无误，则执行N次规范化操作，每次随机选择1.1~1.4四种之中的一种规范化方法。将每次执行规范化的数据集，输出到文件，文件名格式为：

<数据集名>-<规范化方法>-<类标>-<文件产生日期>.txt，如：

Yale\_15cl\_1024f\_165r-Z\_score\_scaling-1-2016\_11\_9\_14\_20\_15.txt,

Yale\_15cl\_1024f\_165r-Z\_score\_scaling-2-2016\_11\_9-14\_20\_18.txt,

…

Yale\_15cl\_1024f\_165r-0\_1\_scaling-1-2016\_11\_9-14\_22\_11.txt,

…

2. 程序设计要求：

源代码应尽量设计成模块化，以便合理组织程序，达到较好的模块化程序设计水平。

1. 提交所有源代码和设计说明。
2. 设计说明

（至少覆盖：UML类图+说明、时序图（每个功能一张时序图）+说明、数据结构设计说明）。设计说明需回答做出关键的设计决定的理由或依据是什么，为什么要这样设计，并且回答以下问题：

1. 为什么要定义这样一个类，这个类名命名是否体现单一责任原则、名词化、拟人化、功能完整性、权力下放、对象自治性，怎么体现？
2. 每个类的公有方法是否控制在很少的范围（如公有方法不超过10个），公有方法越少，则越方便用户使用，同时越能体现类的单一责任原则。每私有方法是否都相应地支撑公有方法。
3. 每个方法体的复杂性是否何控制的，包括每个方法体的代码行数是否控制在200行以内，循环深度是否控制在2层以内？
4. 对象之间的交互（消息传递）是否过于复杂，对象之间的交互包括对象之间是如何建立引用关联的（指针关联的）？，传递的参数是否过多？，传递的参数是否包括不必要的、每次都需要传递的参数？
5. 面向对象编程是有状态编程，每个对象是否都保留有它应该管理的数据成员（状态），它管理的数据成员不多（即具有单一责任属性）、也不少（即保证对象的数据完整性）。
6. 每个类的方法是否都尽可能只依赖该类的成员变量？，还是需要结合内部成员变量与外部传入的参数？为每个类统计一下：该类有多少方法需要结合传递notrivial参数才能正常工作？当这种类型的方法占该类总方法数 > 20%时，这个类的方法是否应该合并到调用方所在的类中成为私有方法，为什么？
7. 为你所写的每个类，统计一下：每个类的代码行数、共有多少方法、共有多少属性、共有多少外部传入的参数用于完成这个类的功能、每个方法有多少行代码。作出一张列表：|类名 ｜ 方法数量 ｜ 属性数量 ｜ 最大方法体的代码行数 ｜ 最小方法体的代码行数 ｜ 类中所有方法代码行数分布情况（平均值、中位数、标准差） ｜ 外部传入的参数数量 ｜ ，并分析这张表中，哪些类的责任过载？哪些类的责任过轻？如何分析？、如何改进？
8. 统计类与类之间的消息交互量，做类似协方差矩阵：

类1 ｜ 类2 ｜ … | 类n

类1 类1私有函数调用次数

类2 类1与类2消息调用次数

.

.

.

类n

利用上表，分析类之间的协作密集性，从而决定是否需要合并类，还是拆分类。

1. 可能的改进说明
2. 实验总结

（此处可以陈述自己在完成该实验时遇到的困难和弯路，以及可能的进步方向等。）

|  |
| --- |
| 指导教师批阅意见：  成绩评定：  指导教师签字：  年 月 日 |
| 备注： |

注：1、报告内的项目或内容设置，可根据实际情况加以调整和补充。

2、教师批改学生实验报告时间应在学生提交实验报告时间后10日内。