**《数据库系统课程设计》报告**

|  |  |
| --- | --- |
| 报告题目 | 数据库系统课程设计 |
| 专业 | 软件测试 |
| 小组成员 | 学号：201545036 姓名： 熊楚原  学号：201592109 姓名： 周 昊 |
| 任课教师 | 朱 明 |
| 完成日期 | 2018/7/24 |

大连理工大学软件学院

**课程设计参数信息**

|  |  |  |
| --- | --- | --- |
| **代码量** | **设计图的个数** | **数据库表的个数**  **CoreData/后台** |
| **500** |  | **0** |

**课程评价结果**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **1级评价指标** | **2级评价指标** | **评价方式** | **评价分值** | **评价结果** |
| **4.研究：**能够基于科学原理并采用科学方法对复杂工程问题进行研究，包括设计实验、分析与解释数据、并通过信息综合得到合理有效的结论。 | （初级）③ 通过信息综合得到合理有效的结论的能力，包括从实验中呈现资料、解释资料或结果，并总结结论及给予建议，形成报告 | 报告、答辩 | 20 |  |
| **5.使用现代工具：**能够针对复杂工程问题，开发、选择与使用恰当的技术、资源、现代工程工具和信息技术工具，包括对复杂工程问题的预测与模拟，并能够理解其局限性。 | （高级）③ 对复杂工程问题进行预测与模拟，并能够理解其局限性 | 报告、答辩 | 50 |  |
| **9.个人和团队：**能够在多学科背景下的团队中承担个体、团队成员以及负责人的角色。 | （中级）② 在设计与开发解决方案过程中明确角色，承担责任 | 报告、答辩 | 30 |  |

目 录

摘要 1

1 功能概述 2

1.1 系统功能 2

1.2 运行环境 2

2 概要分析 2

3 详细设与实现 3

4 成果展示 10

5 系统测试 15

6 课程设计总结 15

## 摘要

随着经济的快速发展，人才之间的竞争变得日益激烈，人力资源也顺势成为了企业的核心竞争力。因此，如何有效的进行人事管理，维系控制好团队的稳定，降低成本，准确有效预测员工的离职倾向，并通过最优的策略实现留住核心人才的目的便成为企业的重要任务。

随着大数据时代的到来，受益于可用计算力与数据量的飞跃提升，深度学习得到了极大的发展。神经网络是目前最接近人类思维方式的一种人工智能方法，它可以通过模拟人脑结构来组织构造网络结构，能够发现高维数据的复杂结构特点，在很多地方的性能都胜过其他机器学习方法。强化学习通过试错的方式不断地与环境进行交互，使动作从环境中获得最大的累积奖赏值，进而发现最优的行为策略。

鉴于以上方法在各个领域的非凡表现，所以我们将其应用到企业人事管理中来进行员工的离职预测。我们首先从员工的基本信息以及打卡行为数据入手，针对企业员工的考勤数据，通过进行特征工程的工作使其可计算，并利用基于多层感知机的深度学习模型对数据进行建模处理，最终在员工的相应数据中预测出其离职的风险，并将结果进行可视化处理分析，紧接着利用基于强化学习的SARSA算法给出降低员工离职率的最优推荐策略，从而实现基于大数据分析的精细化个性化企业人事管理。

**关键词：离职预测；神经网络；最优策略；强化学习**

1 功能概述

## 1.1 系统功能

**本系统已实现预测员工的离职率，并且基于此功能自动化生成一系列离职解决方法的策略。**

* **自动化识别表格**
* **特征工程封装化**
* **离职率预测**
* **离职策略自动生成**

## 1.2 运行环境

|  |  |  |
| --- | --- | --- |
| 项目开发硬件要求 | CPU | Intel Core i5 |
| RAM | 8 GB |
| 硬盘 | 500G |
| 操作系统 | macOS Sierra 10.12 |
| 项目开发软件要求 | 开发平台 | TensorFlow-1.4.1 |
| 开发语言 | Python-3.6 |
| 数据库 | 无 |

2 概要分析

1. 概述

关键人才的流失对于企业来说是有重大负面影响作用的。因此，如何留住企业关键人才不可避免的成为了大多数人力资源部门的关键任务。所以，提前洞悉员工离职概率，对其是否具有离职倾向进行有效预测，以便企业人力资源部门对即将离职的关键人才做出快速反应，给出针对性应对措施，从而为企业留下对于发展有重大作用的关键人才，是具有重要现实意义的研究课题。而我们要解决的问题便是：针对企业员工的可能离职现象，提出一种离职倾向预测算法，帮助企业提前获悉员工的离职动态，以便留下核心人才。随着大数据和云计算的快速发展，与传统算法不同的是，企业所有的海量的员工考勤数据在神经网络里具有了用武之地。考勤数据直接表现了一个员工的工作状态，其中必定隐含着某种规律可以用来反映员工是否具有离职倾向。所以，我们尝试对企业提供的员工考勤数据进行深入挖掘，分析提取其中的隐形特征来对员工是否离职进行预测。对于处理信息量大的数据集，神经网络技术近年来在多个领域都取得了较好的结果。

（2）总体设计

3 详细设与实现

1. 特征工程

数据收集

数据来源：IBM Watson 大数据分析平台

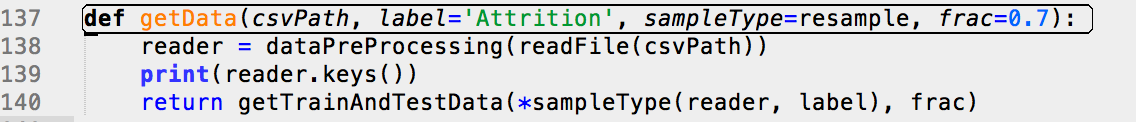
特征选择



特征降维

31维数据降到27维

数据预处理与采样



1. 开发环境

物理设备为MacOS，开发环境为TensorFlow-1.4.1。

（3）用到的知识

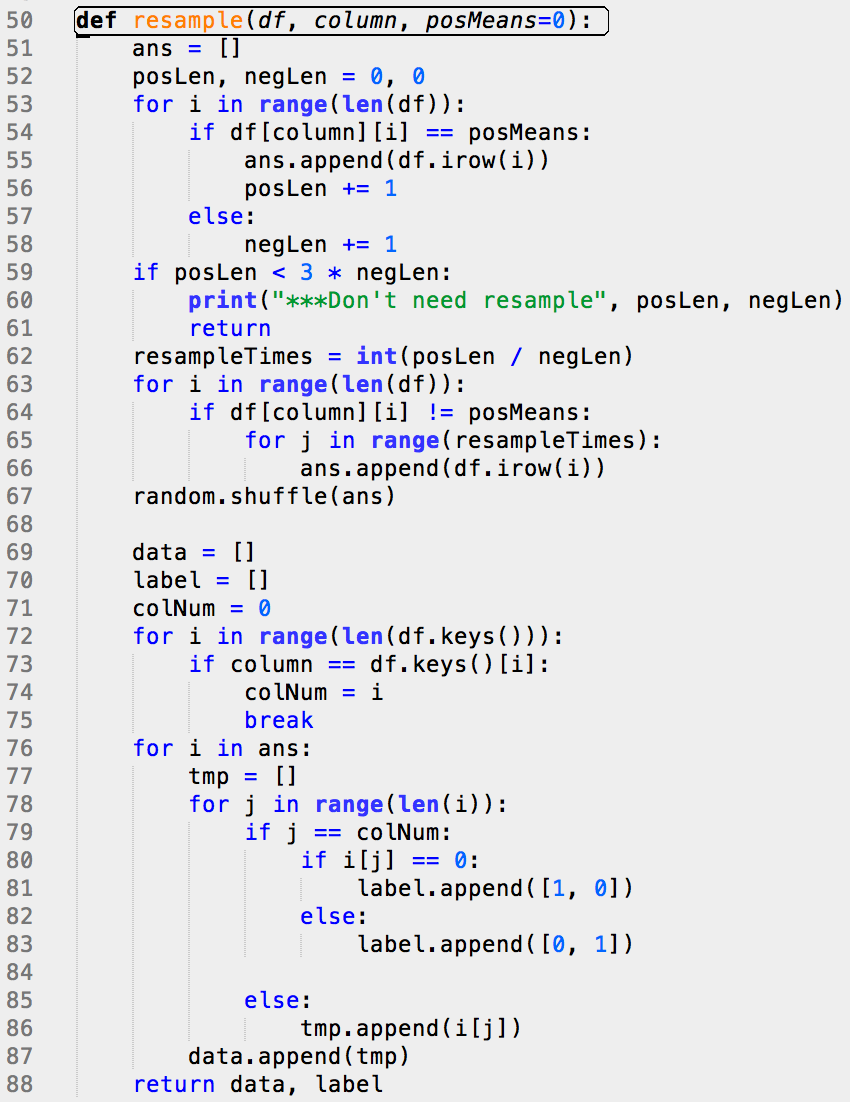
特征工程：统计学与大数据分析

深度学习：bp神经网络

强化学习：Sarsa算法

（4）特征工程核心代码

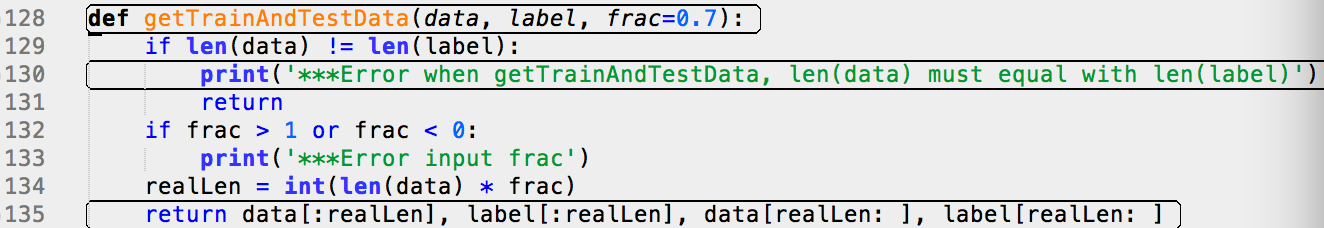
重采样：



欠采样：



抽样：

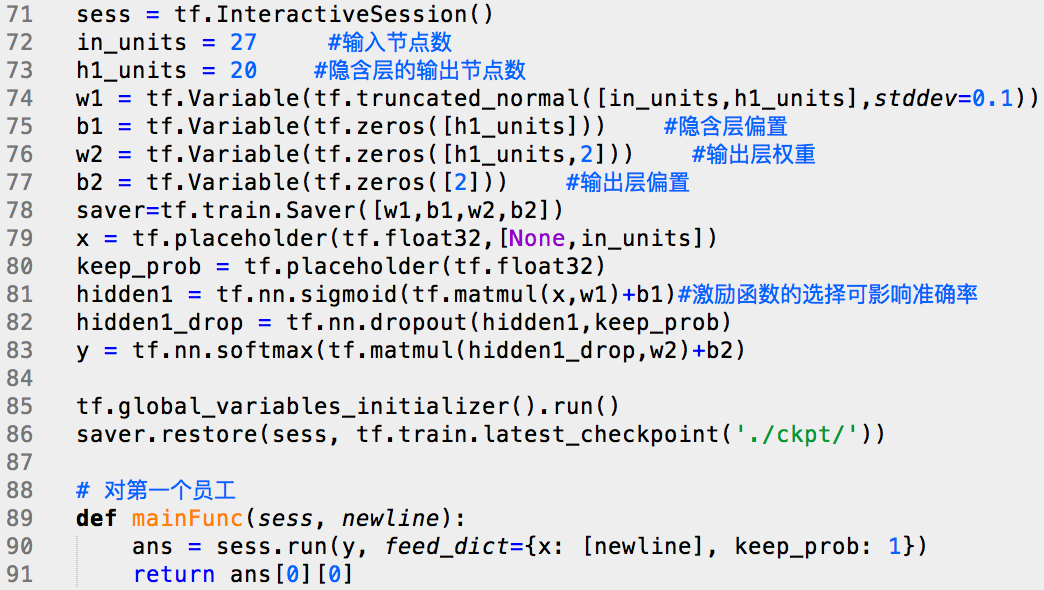


（5）离职率预测

核心训练代码：



封装mainFunc：



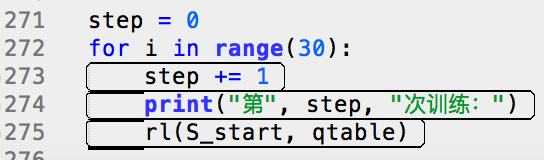
（6）元动作定义（成本相同的动作）



（7）强化学习一次训练过程

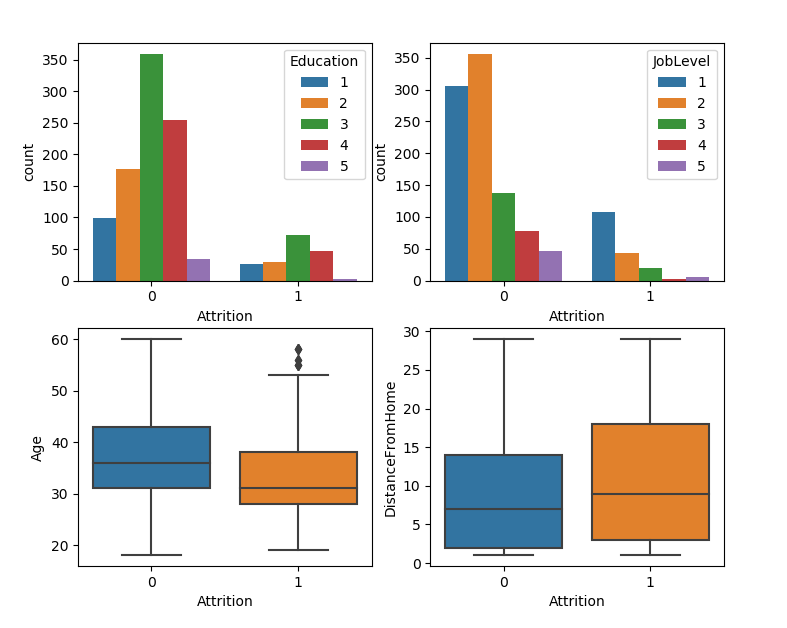


（7）强化学习30次训练过程

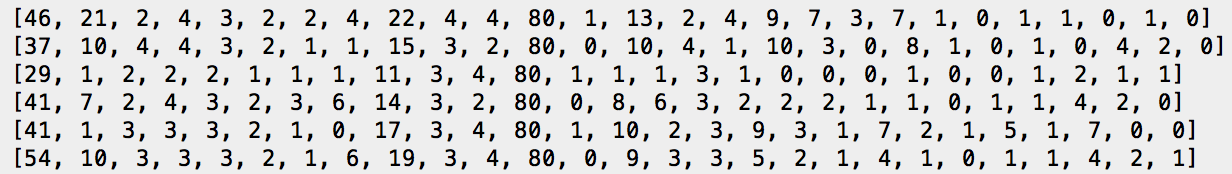


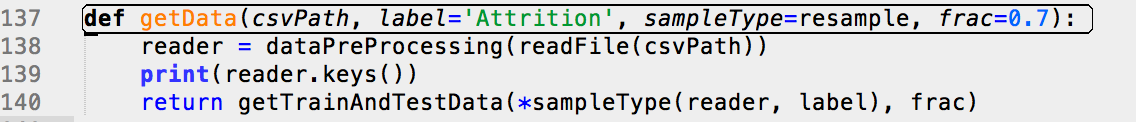
1. 成果展示
2. 特征工程处理结果

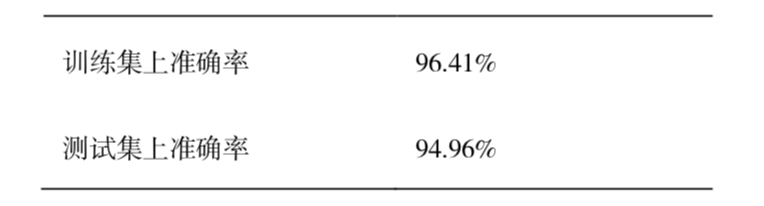
特征相关性预分析：



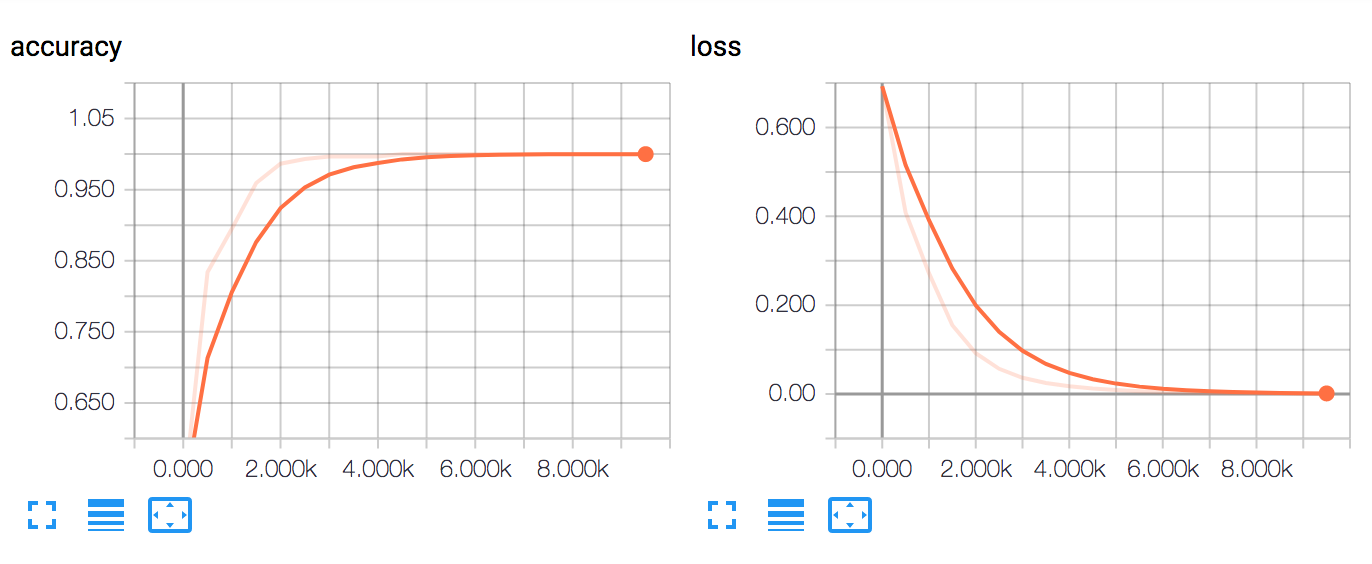
最终处理成的矩阵和采样函数：



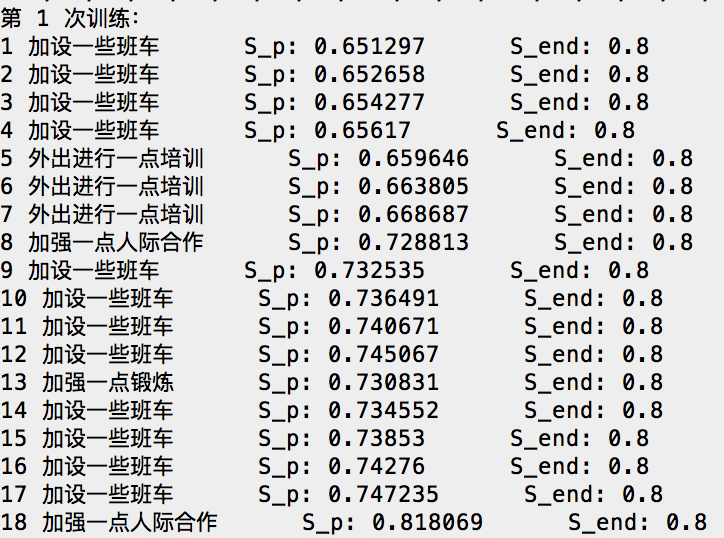


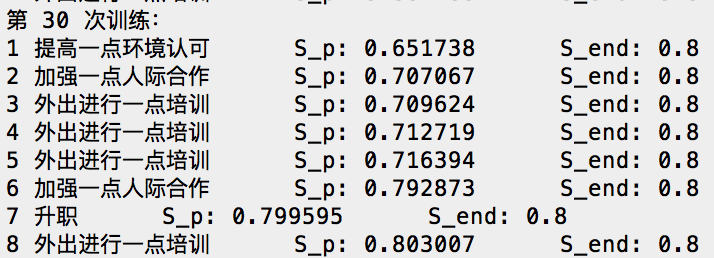
（2）离职率预测准确率

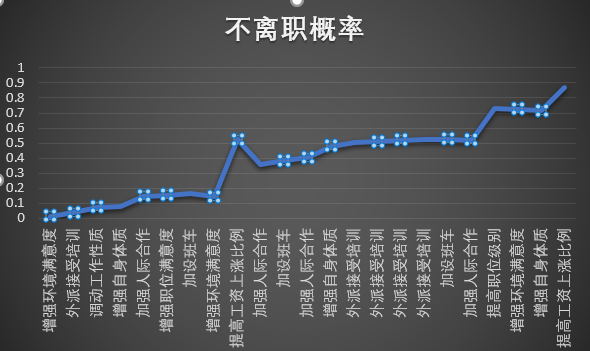
TensorBoard上观察

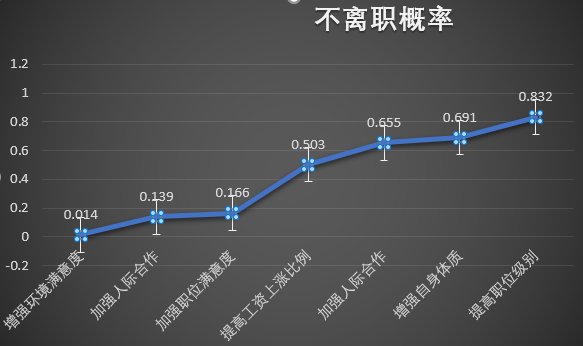


（3）强化学习训练成果









1. 系统测试

（1）测试方法

测试方法主要为黑盒测试，少量白盒测试。代码量较少。测试相对简单

1. 课程设计总结

1.预测准确率高：在训练集和测试集上准确率都达到了至少90%,比其他参观统计器方法准确率大大的提升

2.潜力大：在实验性数据上可以达到很好的效果，进一步研究可以很大的降低企业的人力成本

3.使用方便：直接调用函数输入数据表指定标签即可批量得到员工离职率，很方便

4.切合时代：该项目追求准确率高，自动化，实用性强等特点非常适用于当今时代