**总体设计方案**

**编制:雷邓君**

**项目名称:公平价二手车估值升级**

**本文件修订历史信息:**

|  |  |  |  |
| --- | --- | --- | --- |
| **更新日期** | **修订者** | **批准人** | **主要修订摘要** |
| 2018.10.31 | 雷邓君 | 万可文 | 新编写文件 |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

# 1.总体目标

## 1.1总体目标

估值系统:能通过品牌,车系,款型,公里数,城市,上牌时间特征预测该车当前的车商收购价,车商真实售价,车商建议标价,个人交易价,最低成交价,认证二手车价,4S店置换价,车商批发价,平均买车价,平均卖车价十类价格.每种价格分四个车况,[优秀,良好,一般,较差].

残值系统:能通过品牌,车系,款型,公里数,城市,上牌时间特征预测该车当前的车商收购价,车商真实售价,个人交易价未来12个月每个月的残值和未来三年每年的残值,以及过去6个月这三类价格的历史走势.

# 2.需求分析及功能设计

## 2.1子系统划分

估值大系统分成三部分:爬虫系统,款型匹配系统和估值算法系统。

#### 2.1.1 爬虫系统

功能介绍:抓取两部分内容,一是汽车之家的款型库作为我方标准库的补充和新款车4S店的实际售车价,二是抓取瓜子,人人车,优信三平台的车源数据作为估值算法系统的训练数据.详细设计请参见吴辛成编写的《爬虫系统设计》

#### 2.1.2 款型匹配系统

功能介绍:通过车源信息的文字标题,预测其品牌,车系,款型与我方标准库的对应关系.

例如文字标题:奥迪A8L 2013款 A8L 40 hybrspecid(进口)

预测品牌:奥迪

预测车系:奥迪A8L

预测款型:2013款 A8L 40 hybrspecid(进口)

#### 2.1.3 估值算法系统

功能介绍:将爬虫抓取的车源数据作为训练数据进行训练,并根据款型,上牌年份,公里数,城市预测二手车的十类价格.

## 2.2系统流程分析

1. 爬虫抓取汽车之家款型库
2. 将我方款型库与汽车之家款型库进行匹配并最终生成我方标准库
3. 将我方标准库的款型文字描述作为训练数据，使用深度神经循环网络lstm算法进行训练并最终生成品牌预测,车系预测,款型预测三大模型.
4. 爬虫抓取瓜子,人人车,优信三平台车源数据。
5. 用3生成的模型预测4抓取的车源数据与我方标准库的对应关系
6. 将5匹配上的数据进行数据探索,特征工程,生成训练数据并使用xgboost算法最终训练生成估值模型.

# 3.关键算法介绍

## 3.1深度神经循环网络lstm介绍

请参见网址:https://blog.csdn.net/yoyofu007/article/details/80361422

## 3.2 xgboost算法介绍

请参见网址:http://blog.itpub.net/31542119/viewspace-2199549/

# 4.技术选型

语言:python3 使用原因:简练,易于阅读安装扩展,交互性,多范式,多平台,多框架支持

框架:tensorflow 使用原因:TensorFlow是由Google公司开发的用于实现机器学习的一种架构,是大公司背景,文档完善,使用人多.介绍文档:https://blog.csdn.net/qq\_34580082/article/details/79633535

部署:微服务+ubuntu+docker 使用原因:服务简单,只关注一个业务功能,每个微服务可由不同团队开发,微服务是松散耦合的,可用不同的变成语言与工具开发.介绍文档:https://blog.csdn.net/wuxiaobingandbob/article/details/78642020?locationNum=1&fps=1

# 5.应用安装和验收

## 5.1 安装

请参见《估值模块使用说明》

## 5.2 验收

交付python估值调用包