第十届中国大学生服务外包创新创业大赛



详 细 方 案

项目名称：运用文本相似度实现（证券）智能客服

命题方向：人工智能

题目类别：计算类

目录

[一、 目标与服务模型 2](#_Toc4427921)

[1.1 项目背景 2](#_Toc4427922)

[1.1.1整体背景 2](#_Toc4427923)

[1.1.2公司背景 2](#_Toc4427924)

[1.1.3 业务背景 3](#_Toc4427925)

[1.1.4 团队背景 3](#_Toc4427926)

[1.2 目标价值 3](#_Toc4427927)

[1.2.1 项目目标 3](#_Toc4427928)

[1.2.2 产品功能 4](#_Toc4427929)

[1.3产品内容 5](#_Toc4427930)

[1.4 产品亮点 5](#_Toc4427931)

[1.5 用户思路 6](#_Toc4427932)

[1.5.1 客户 6](#_Toc4427933)

[1.5.2管理者 6](#_Toc4427934)

[二、 过程管理与业务模型 6](#_Toc4427935)

[2.1 过程管理 6](#_Toc4427936)

[2.2 业务分析 7](#_Toc4427937)

[2.2.1 系统目标 7](#_Toc4427938)

[2.2.2 用户需求 7](#_Toc4427939)

[三、 技术路线及实施方案 7](#_Toc4427940)

[3.1 总体介绍 7](#_Toc4427941)

[3.2 开发工具及环境 8](#_Toc4427942)

[3.3 处理流程 9](#_Toc4427943)

[3.3.1 文本处理流程 9](#_Toc4427944)

[3.3.2 文本匹配流程 10](#_Toc4427945)

[3.3.3 技术指标验证 11](#_Toc4427946)

[3.4 技术路线 11](#_Toc4427947)

[3.4.1 相似度算法 11](#_Toc4427948)

[3.4.2 相关问题推荐 12](#_Toc4427949)

[四、 可行性分析 13](#_Toc4427950)

[4.1 经济可行性 13](#_Toc4427951)

[4.2 操作可行性 13](#_Toc4427952)

[4.3 技术可行性 13](#_Toc4427953)

[4.4 社会可行性 13](#_Toc4427954)

[4.5 可行性分析结论 13](#_Toc4427955)

1. 目标与服务模型
   1. 项目背景

1.1.1整体背景

AI 时代到来，给互联网金融将会带来全新的体验。客户只需要在手机端输入

感兴趣的金融问题，智能客服就会将正确的答案反馈回来，开户需要哪些材料，费用如何收取……

智能语义识别，不仅给金融提供了智慧服务，还提供了很多的创新空间。

1.1.2公司背景

恒生电子股份有限公司（简称恒生公司）于 1995 年成立于杭州，2003 年在上海证券交易所主板上市（代码 600570），是中国领先的金融软件和网络服务

供应商。恒生聚焦于财富资产管理，致力于为证券、期货、基金、信托、保险、

银行、交易所、私募等机构提供整体的解决方案和服务，为个人投资者提供财富

管理工具。目前，恒生公司拥有约 8000 人的高素质专业队伍，其中研发工程人

员约占 80%，在规划咨询、软件开发实施、技术服务、系统集成及系统维护服务

等方面，具备强劲实力和竞争优势。多年来，恒生以技术服务为核心，凭借多年

金融 IT 建设经验，以及对互联网的深刻洞察和理解，用优质的产品与服务，驱

动金融机构创新发展。

1.1.3 业务背景

客服是企业和客户之间的纽带，承担着直接服务客户的工作。解答业务问题、

处理服务纠纷，加深客户对企业的了解、增进客户与企业之间的感情等等，这些

重要的任务都是客服的日常内容。客服在一定程度上代表着企业的形象。客服工

作可以折射出一个企业的文化修养与综合素质。客服工作的好坏与企业利益是息

息相关的。

然而，客服工作不仅强度大，而且工作内容也比较枯燥乏味。这一工作特性

使得客服人员容易产生负面情绪，进而导致客服人员流动性大、招聘管理难，最

终使得客户体验效果不理想。AI 客服的问世，将极大地缓解企业的这些难题痛

点。因此， AI 客服受到了金融服务业的青睐，被应用于证券、基金等各大金融企业。

1.1.4 团队背景

飞虎大队是一支由三个活力四射的少年组成的团队，队员对技术充满热情，在开始做此项目之前时，团队成员做好了充足的准备工作，满怀信心的选择了这个项目，在项目的开发过程之中，团队成员互帮互助，进益求精

1.2 目标价值

1.2.1 项目目标

根据用户的期望，目标概括如下

* 开发一套能够替代客服功能，为客户的问题咨询提供解答
* 为企业提供智能客服，节省人力成本
* 为客户节约咨询效率。
* 可拓展的短文本相似计算方案。
* 较高准确率的智能回复。
* 支持后台导入标准问答库；
* 支持文本相似度计算功能，提供测验验证；
* 生成问题解答集合，用于开发人员测试算法。

1.2.2 产品功能

本产品的用户角色有客户和管理者

管理者可用功能：导入标准问答库、增删改预设的词库、进行文本匹配测试、查看匹配生成的文档

客户可用功能：在线咨询功能

导入标准问答库

增删改独立词库

管理者端

文本匹配测试

问题表述精确

给出标准解答

用户端

给出相关问题的相似度排序

问题表述模糊

1.3产品内容

1.主体程序

用于进行文本匹配，输出最佳答案。

2.停用词库

用于去除语句中的无关词，例如“啊”，“吗”之类。

3.同义词库

用于进行同义词的转换，例如将“如何”转为“为什么”。

4.关键词库

抓取文本中的专业术语，例如“佣金”，“创业板”。

1.4 产品亮点

* 可拓展性 ：构建了独立的词库，用户可根据需求进行删改。
* 算法创新 ：依据短文本的特性，抛弃了传统的TF-IDF；依据汉语表达的特点，算法中排除了语序的干扰。
* 正确率高 ：专业术词的标注，关键词提取等方法在测试中反映出极高的准确性。
* 容错率高 ：TOP5中无关项较少，在用户提问模糊的情况下可以返回相关性供用户选择。
* 操作便捷 ：界面简洁人性化，一键操作

1.5 用户思路

1.5.1 客户

* 客户需要一个简洁方便的操作界面，来适应个年龄段的客户
* 客户需要系统运行速度快，能迅速给出解答，减少等待时间
* 客户需要智能客服给出准确的解答，或者能够猜想客户所问的问题是什么

1.5.2管理者

* 管理者可以随时删改标准问答库，改变问题给出的标准回答
* 管理者可以对算法进行测试，方便进行维护，

1. 过程管理与业务模型

2.1 过程管理

本项目的需求较为清晰，团队计划打造智能客服系统适应范围更广，并提出如下过程模型。

过程模型图：

综合测试

实现

规格说明

设计

需求分析

需求分析：在需求分析阶段，项目组通过分析赛题、分析已有系统、与出题企

规格说明：产生完整描述系统功能、性能的《需求规格说明书》，为项目设计人员和项目开发人员进行详细设计和编程提供基础

设计 ：制作快速原型，确认最终期望得到的产品的外观和功能。进行技术选型，进行技术分析，制定开发计划，设计出相关词库和相似度算法

实现：向开发人员分配任务，由开发人员通过编码实现整套系统

综合测试 ：对项目进行测试，根据给出的算法测试要求，得出F1值

2.2 业务分析

2.2.1 系统目标

AI客服系统的目的是为了更好的满足客户，弥补人工客服的不足，为企业带来更好的形象和经济效益。

2.2.2 用户需求

1. 提供一个可扩展的短文本相似度计算方案；

2.在相似度大于 0.8 的条件下，top1 的正确率大于等于 0.8，召回率大于

等于 0.75， 值尽可能大；

3.在相似度大于 0.6 的条件下，top5 中无关内容尽可能少。

1. 技术路线及实施方案

3.1 总体介绍

本项目名为运用文本相似度实现（证券）智能客服，运用python进行开发，构建有独立的词库进行文本的自动修改。计算文本相似度并导出成txt格式，便于管理者进行测试。

3.2 开发工具及环境

开发工具：pycharm；tkinker；python IDE；Eclipse；

开发技术：python；

软件环境：

（1）开发环境：

windows7及以上版本

（2）相关软件：

Excel2003以上版本

硬件环境：

不限

3.3 处理流程

3.3.1 文本处理流程

用户输入文本

文本预处理完成

替换同义词

去除停用词

文本匹配算法

文本相似度算法

文本匹配

标识专业词

分割关键词

给出解答

3.3.2 文本匹配流程

例：专业词

创业板

抓取专业词

预处理后的文本

得到专业词有关项

创业板权限转签如何操作？

创业板权限首次开通？

创业板如何开通？

删除停用词

词袋模型

（使用结巴分词）

替换关键词

文本余弦相似性计算

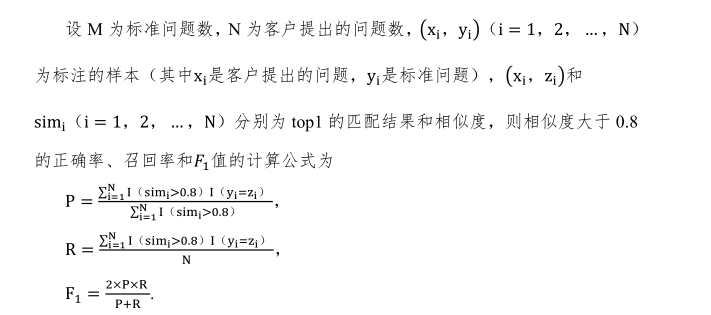
输入文本为样本

计算各标准问与其文本距离

匹配至最相似文本

3.3.3 技术指标验证

根据以下算法，将正确率（P值），召回率（R值），F1值 反馈给管理者



3.4 技术路线

3.4.1 相似度算法

项目了采用Python中的“jieba”（结巴分词），利用其高达 97%以上的分词精度，对文章内容将进行中文分词结巴分词。虽然能获取文章的关键字，但关键字中会存在一定的无意义的词语，例如“啊”，“吗”。因此采用了设置停用词词库的方法，优先去除此类无用词。

短文本匹配过程中用户输入的问题中可能存在与已有标准问题的同义文本。例如 “为什么”与“为何”，“怎么解决”可以归为同义词，但在相似度算法上会造成影响。设置了同义词库，对同义词进行转换来提升文本相似度的准确性。

对于金融企业而言，一些金融领域术词在用户输入的文本中将占有更高的权重，优先进行匹配，这一操作将由专业词库及其配套算法进行完成。

提取完关键词并进行处理后，接下来需构建词袋空间 VSM(vector space model)。

首先统计所有文档的词集合，对每个文档都将构建一个向量，向量的值是词语在本文档

中出现的次数。如下列短文本：

|  |  |  |  |
| --- | --- | --- | --- |
| 我 | 今天 | 真 | 开心 |
| 我 | 今天 | 开学 | 了 |

定义向量集（我，今天，真，开心，开学，了）

原来两句话转换成数字向量 ：

a：（1.1.1.1.0.0）

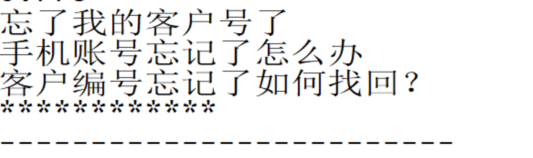
b：（1.1.0.0.1.1）

利用数学模型 计算相似度

其中 a 为标准文本 b为输入文本

3.4.2 相关问题推荐

基于文本相似度进行的相关问题排序。程序首先测算出输入文本与标准文本之间的相似度，其中专业词占有更大权重。然后将有关问题列出，供给用户选择。



1. 可行性分析

4.1 经济可行性

支出：项目开发成本估计数值在客户预期范围内。

收益：系统能智能地处理客户的问题，给予标准的回答，为企业大大节省了人力成本。给客户提供了更加智能、更加人性化的交互体验，显著提高了客服效率。同时可拓展的信息库可以导出，用于其他类型的智能客服。

投资回收时间大约为半年

4.2 操作可行性

系统界面简洁，操作简单，只需导入语料包即可运行，拓展性强，仅需简单教程即可熟练操作。

4.3 技术可行性

本程序通过python语言编成，逻辑并不复杂，数据保存修改都通过文件进行，技术难度简单，可拓展性强，具有良好的稳定性与安全性，使程序能够稳定运行。相似度计算方法简单明了，容错率高，能够胜任普通的短文本相似度计算需求。

4.4 社会可行性

本系统自主研制开发，开发将不会侵犯任何个人、集体、国家的利益，也不会违反国家的政策与法律。

4.5 可行性分析结论

本产品经济效益较高，拓展性强，能够智能地处理客户的问题，给予标准的回答，为企业大大节省了人力成本。给客户提供了更加智能、更加人性化的交互体验，显著提高了客服效率。本产品促使金融客服向着更加智能的方式发展，以跟上智能时代的步伐。本项目具备可行性。