



黑马程序员线上品牌

# 金融行业动态方向评估项目

---

一样的教育，不一样的品质



# 目录

Contents

1. 项目背景介绍
2. 项目任务与方法介绍
3. Few-shot、Zero-shot思想回顾
4. 项目环境配置

# 01 项目背景介绍



当前金融领域数据大量激增, 如何从繁杂的数据中获取有效的信息, 进而帮助投资者或者研究者减少决策失误带来的损失, 成为金融数据分析方法研究的热门话题。

随着科技的进步, 人工智能技术在各行业中的应用越来越广泛, 而金融领域也不例外。人工智能技术的应用可以为金融企业提供更高效、精准的服务, 也可以帮助投资者更好的地进行投资决策。

## AI在金融行业的应用

01

### 风险评估

通过AI技术识别出不同金融风险事件类型、反欺诈行为、信用评级低等风险，提高贷款的准确性和风险控制能力。

02

### 投资决策

通过AI技术对历史数据、财务报表等信息分析，为投资者提供精准的投资决策支持。

03

### 客户服务

通过AI技术，实现智能客服等功能，进而为客户提供便捷而又准确的答案。

## I AI应用方式

### 传统应用方式

需要专业的AI技术加持如：NLP、CV、机器学习等知识

专业的算法人员去分析数据、训练模型从实现预测分析

非专业人员来说要想实现上述应用，可谓是举步维艰。

VS

### 大模型应用方式

伴随着ChatGPT等大模型问世，使得非专业人员实现上述应用成为可能

不需要特别专业的算法知识，就可以利用大模型来实现金融领域的应用。

设计合理的Prompt来挖掘大模型的涌现能力。

因此，本项目主要基于大模型来直接实现在金融领域相关任务的应用。重点在于如何对大模型设计prompt，从而激发大模型的"涌现能力"，进而给出准确的答案。



# 02

## 项目任务与方法介绍

## 项目任务

### 业务场景

#### 金融文本分类

如：帮助机构识别和评估不同类型的风险，包括信用风险、市场风险、操作风险等。

#### 金融文本信息抽取

如：自动识别和提取金融文本中的关键信息，如实体名称、日期、金额等，可以大大提高数据处理的效率，减少人工处理的工作量。

#### 金融文本匹配

如：信息检索和过滤，匹配相关的金融信息文本，提供准确的搜索结果。

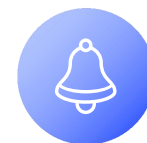


## I 方法选择



### 模型

ChatGLM-6B



### Incontext-Learning

基于Few-shot/zero-shot  
等方法，设计Prompt

注意：本项目主要以金融领域的文本进行分析，但是该思想同样适用于其他场景数据。

# 03

## Few-shot、Zero-shot思想回顾

## Zero-Shot

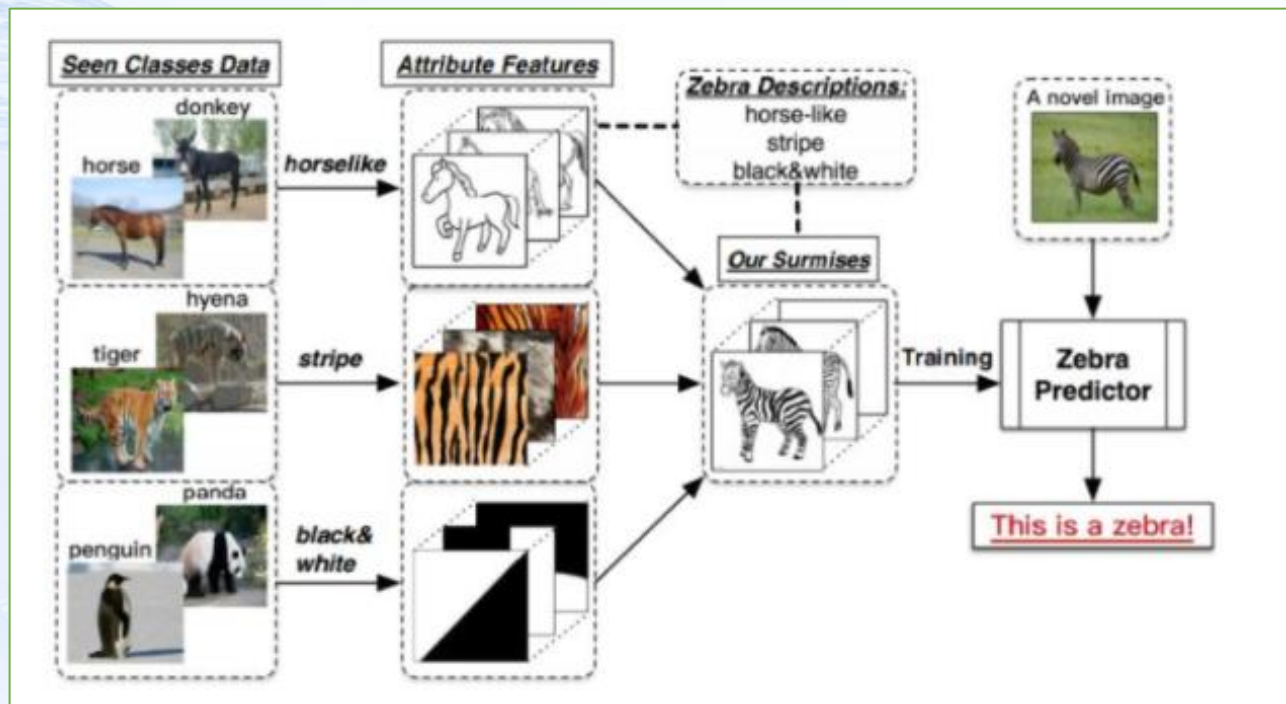
Zero-shot (Zero-shot Learning) 是指在训练阶段不存在与测试阶段完全相同的类别，但是模型可以使用训练过的知识来推广到测试集中的新类别上。这种能力被称为“零样本”学习，因为模型在训练时从未见过测试集中的新类别。

### 举例说明：

有一天，小明和爸爸一块去动物园，看到了马，然后爸爸告诉他，这就是马；之后，又看到了老虎，告诉他：“看，这种身上有条纹的动物就是老虎。”；最后，又带他去看了熊猫，对他说：“你看这熊猫是黑白色的。”然后，爸爸给小明提了个问题，让他在动物园里找一种他从未见过的动物，叫斑马，并告诉了小明有关于斑马的信息：“斑马有着马的轮廓，身上有像老虎一样的条纹，而且它像熊猫一样是黑白色的。”最后，小明根据爸爸的提示，在动物园里找到了斑马（意料之中的结局。。。)

## Zero-Shot

上述例子中展示了一个人类的推理过程，就是利用过去的知识（马，老虎，熊猫和斑马的描述），在脑海中推理出新对象的具体形态，从而能对新对象进行辨认。（如下图所示）ZSL就是希望能够模仿人类的这个推理过程，使得计算机具有识别新事物的能力。



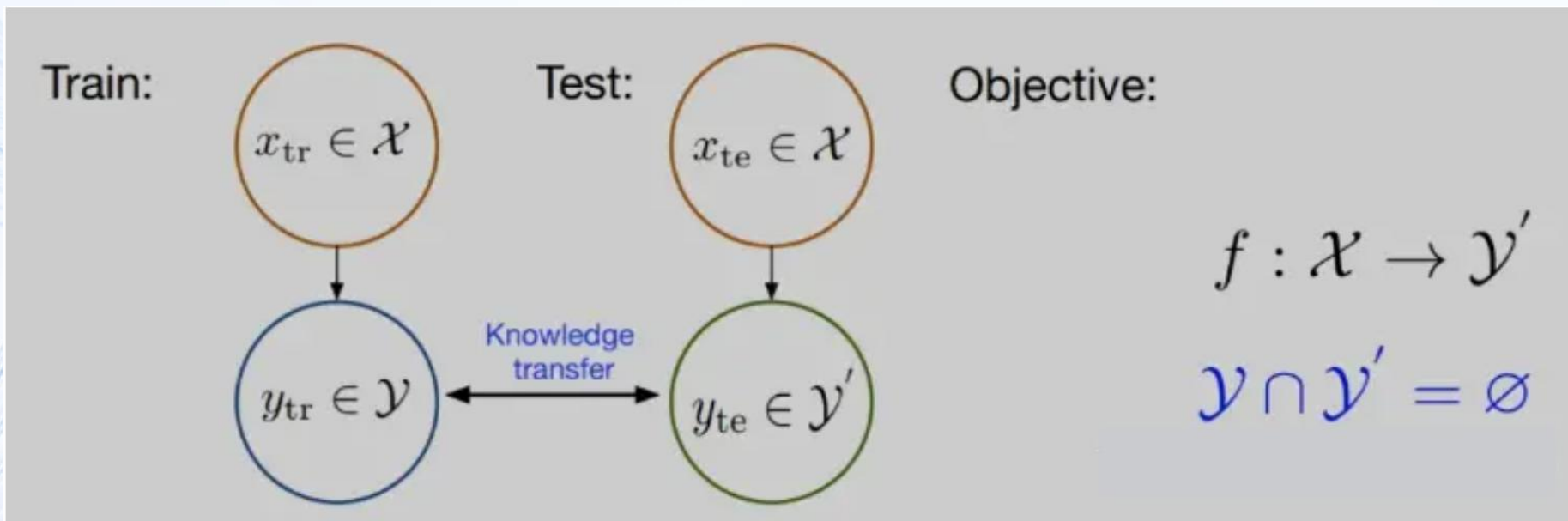
### 与传统算法区别

在传统的图像识别算法中，要想让计算机认出斑马，往往需要给计算机投喂足够量的斑马图像才有做到。而且，训练出来的分类器，往往无法识别它没有见过的其他类别的图像。

## Zero-Shot

模型在zero-shot情况下如果需要准确识别出斑马，需要知道的信息是马的样本、老虎的样本、熊猫的样本和样本的标签，以及关于前三种动物和斑马的描述。

训练和预测基本流程：





## Few-Shot

Few-shot学习 (Few-shot Learning) 是指少样本学习，当模型在学习了一定类别的大量数据后，对于新的类别，只需要少量的样本就能快速学习(目的：让机器自己学会学习)。



Query: 送入模型的问题，依次与support对比，发现和水獭最相似。因此，query应该是水獭。

Support set: 带标签的术语。规模很小，每一类有一张/几张图象。在做预测的时候，提取额外的信息。

Few-shot学习的核心：用足够大的模型，来训练一个大模型。训练的目的，不是让模型单纯识别大象/老虎等。而是让模型学会事物的异同，然后根据support set，就可以知道query是什么，尽管训练集中没有水獭类别。



# 04

## 项目环境配置

## I 环境准备

python版本:

```
python>=3.8
```

依赖包requirements:

```
protobuf>=3.19.5,<3.20.1
```

```
transformers>=4.26.1
```

```
icetk
```

```
cpm_kernels
```

```
streamlit==1.17.0
```



黑马程序员线上品牌



扫码关注博学谷微信公众号

