# 机器学习工程师纳米学位毕业项目

## **《[句子相似度匹配](https://github.com/udacity/cn-machine-learning/tree/master/quora-question-duplicate" \t "https://classroom.udacity.com/nanodegrees/nd009-cn-advanced/parts/3a867cbf-2876-427e-8e3f-bed2ac7111f5/modules/06b6f1e8-b65f-42a9-ab09-799a23f069ad/lessons/72c558d9-a4b3-4c20-8f20-8023afad4cb6/concepts/_blank)》**开题报告

陈宁

2019年6月25号

目录

[1. 背景介绍](#_Toc3433_WPSOffice_Level1) [3](#_Toc3433_WPSOffice_Level1)

[2. 问题描述](#_Toc32317_WPSOffice_Level1) [3](#_Toc32317_WPSOffice_Level1)

[3. 数据集和输入](#_Toc22259_WPSOffice_Level1) [3](#_Toc22259_WPSOffice_Level1)

# 背景介绍

[引用](http://www.sohu.com/a/100532291_157627)

让机器来理解人类语言一直都是人工智能的梦想，最先从词到短语到句子， 再到段落到整篇文章。所有的方式都是将字符串转换为向量，最终从数学的 角度来理解语义。

# 问题描述

[引用](https://www.cnblogs.com/huilixieqi/p/6493089.html)

本项目中提供了已经配对好的句子对，需要用已有的数据集进行训练，最终可以预测两个句子的相似性。这属于监督学习类型。可以用编辑距离类似的方法来解决。可以把每一个句子看成是一个向量，再求两个向量之间的关系。这种关系在每一个句子对中几乎都存在。把这种关系保存下来，当有新的一对未知关系的句子对需要检测时，那么就可以先根据数据模型算出他们的关系，再和训练数据的句子对的关系对比，可以得到新的句子对的相似性。

# 数据集和输入