

# POI 名称生成

Cicero\_Ning

宁星星

认知智能实验室&算法工程师

Percent

中国-武汉

shenzaimin0@gmail.com

镇诗奇

认知智能实验室&算法工程师

Percent

中国-武汉

13476111882@163.com

## 团队简介

队伍成员来自百分点认知智能实验室。

队长：宁星星

毕业于华中科技大学，热爱探索、乐于分享且擅长解决实际问题。

技术方向：语义解析、机器问答以及文本校对算法等。多次在 CCKS、NLPCC 等会议评测算法大赛中获奖。

队员：镇诗奇

中国地质大学（武汉）计算机硕士

技术方向：智能问答、语义解析和关系抽取等，曾获得 CCKS2021 图谱问答第三名和国家电网运行信息智能检索 Text2SQL 比赛三等奖。

## 摘要

本次《POI 名称生成》赛题要求是“给定图像和挂牌检测结果和 OCR 识别结果设计算法，生成图像中包含的完整 POI 名称”。

分析数据集得知，比赛数据集来自高德地图的真实应用场景数据，训练集包含 2.1W 条 POI 信息，每条都包涵原始图片、OCR 识别文本以及挂牌和文本行的轮廓数据。怎样利用从这些不同形式的数据信息中获取真正有用的信息，成为了完成比赛任务的关键。

本次比赛我们采用结合图像位置信息和文本信息的 text2text 生成算法。文本生成模型采用端到端的 Text2Text 预训练模型 T5 框架，在解码过程中，采用 sampling 解码策

略。在后处理时，基于原始 block 的 OCR 识别结果进行匹配校正，得到最终 poi name。

最终我们在 A 榜得分 0.7224，B 榜得分 0.7158，获得第五名的成绩

## 关键词

POI，文本生成，OCR，sampling

## 1 赛题分析

本赛题属于文本生成领域，即将非语言形式信息作为输入，生成可读的文字表述。文本生成任务大致可分为三类：文本到文本的生成、数据到文本的生成、图像到文本的生成。文本生成技术发展至今，已经有诸多相关的解决方案和学术研究，例如，基于规则、基于规划（Planning-based）以及数据驱动（data-driven）的方法等。

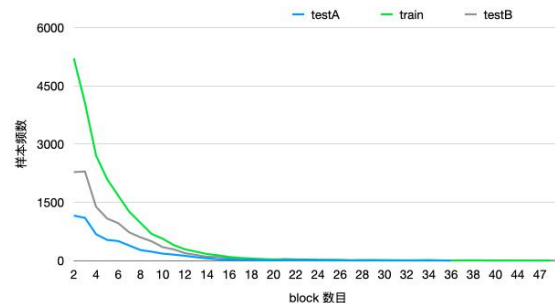


图 1：训练集与测试集在 block num 上的分布

如图 1 所示从数据分布来看，训练集与测试集在输入 block 数目上基本同分布。并且统计真实 poi 名称中包含的 block 数目如下图 2 所示，大部分的 poi 名称都是由 4 个以内的 block 文本框组成。

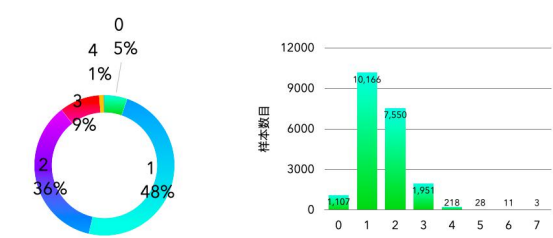


图 2：POI 名称包含的 block 数目与样本频数的关系

2 技术路线

本赛题我们的解决方案为：结合图像位置信息和文本信息的 text2text 生成算法。文本生成模型采用端到端的 Text2Text 预训练模型 T5 框架，在解码过程中，采用 sampling 解码策略。然后基于原始 block 的 OCR 识别结果进行匹配校正，得到最终 poi 名称

2.1 总体架构

本方案的总体算法架构图如图 3 所示。架构图中包含两部分，分别是数据处理部分和模型算法部分。

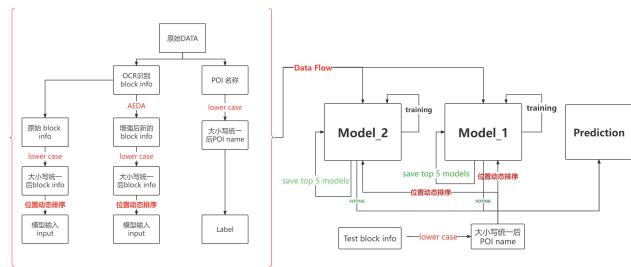


图 3：方案总体架构图

2.2 数据处理

数据处理的关键环节如图 4 所示，主要以下几步：

①对特定样本采用数据增强技术：对 block 数目>4 的样本进行重采样

②标签大小写统一（lower\_case）

③基于 block 字符面积进行 block 文本排序，根据比赛提供的 block 信息，结合坐标位置信息，使用相对字符面积比作为排序依据，对文本框进行位置排序，对排序好的文本使用特殊字符进行分隔，组成排序文本信息，作为模型输入

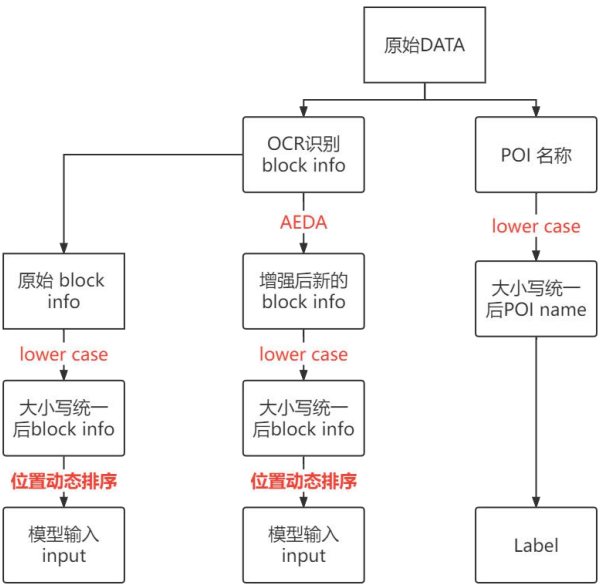


图 4：数据处理流程

2.3 模型算法搭建

本方案算法模型如图 5 所示，主要分为两个模块：Text2Text 文本生成模块和 OCR 信息校正模块。

2.3.1 Text2Text 文本翻译模型

用位置排序模块生成的排序文本，作为生成模型的输入，用 name 作为 lable 进行文本翻译任务的训练 采用完全匹配的评估方式作为评测标准进行模型的评估，选取开发集表现最好的模型作为检查点保存为 model1。同时，对于原始 block 数目大于 4 的样本进行数据增强，训练得到新的模型 model2。

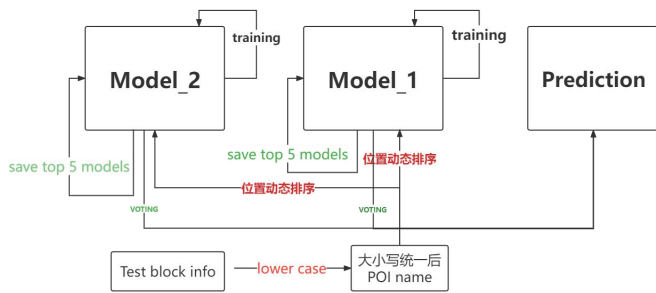


图 5: 算法模型方案

### 2.3.2 基于 OCR 识别的匹配校正模块

对于原始 block 的识别结果进行有限制的排列组合，将文本生成模型的结果与排列组合结果进行字符串文本模糊匹配，取得分最高的结果作为最终生成结果。

## 3 实验成果展示

本赛题我们从图像，坐标以及文本多种信息来源进行了数据挖掘和模型实验，实验内容包括融合多模态信息的文本框分类和回归实验、基于排序文本的生成模型实验、数据增强实验等。

### 3.1 baseline 实验

通过固定 random seed 采用训练集 shuffle 后“二八分”得到 dev 验证集（4207 条）。采用“完全拼接”，

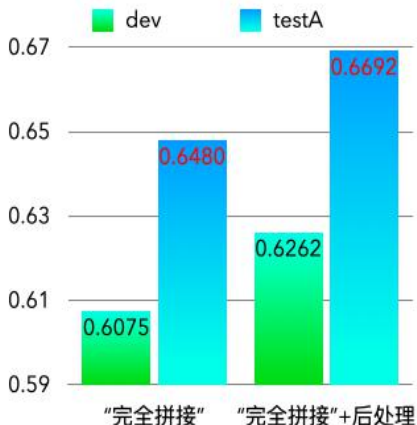


图 6: baseline 实验结果

即样本中所有的 block 文本，以分隔符“|”拼接作为 Text2Text 模型输入，最后对 block 文本全排列与模型生成文本进行模糊匹配后处理。实验结果如下图 6 所示。

从结果可以知道模型结果采用 OCR 文本“矫正”有利于减轻生成模型的“添油加醋”。

### 3.2 多模态信息融合对 block 文本框进行分类

image: 采用 resnet152 对图像信息进行编码

text: 采用 Bert-base 对 block 文本进行编码

分类器: 对图像和文本特征进行简单拼接后，进入全连接层进行二分类，“正样本”为生成 poi 所需文本框，“负样本”为噪音文本框。

通过对比不同 topk 阈值下的全图信息模型和 block 子图信息模型在 testA 数据集上的得分如图 7 所示。

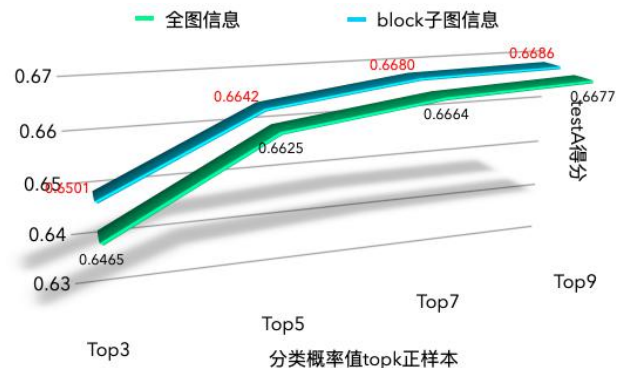


图 7: 不同 topk 阈值下的模型在 testA 数据集上的得分

尽管分类器具有一定的 block 文本框筛选能力，但经过分类筛选后的文本仍然会出现缺少必要 poi block 信息的现象，从而导致模型表现低于 baseline 的表现。

### 3.3 block 字符面积排序实验

通过计算 block 内字符的平均面积作为指标，进行降序排序，拼接后进行生成模型训练。并且根据数据分析统计的 poi 中 block 数目分布，对 block 数目>4 的样本进行上采样后训练。实验结果如图 8 所示。

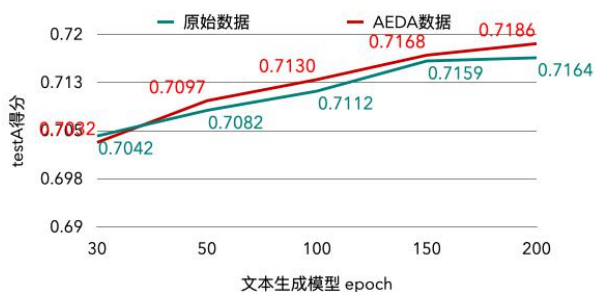


图 8: block 字符面积排序生成实验

实验结果显示，简单的面积排序对 poi 生成结果的提升十分显著。我们最终的 A、B 榜成绩都是在这个模型的基础上集成得到。

#### 4 创新点

创新点 1: End-to-End 方式解决 POI 名称生成任务，执行效率高，无子模型 pipeline 误差传递

创新点 2: 字符面积排序模块，简单直接的输入构造，提升文本生成模型的收敛效果

创新点 3: OCR 识别结果校正，通过字符串模糊匹配，补齐文本生成模型“添油加醋”的短板

#### 5 方案总结

本次比赛中，我们尝试了使用训练数据中的各类信息进行多模态建模，而最终的结果显示，对效果提升最显著的几个关键点是：

1. 根据广告挂牌设计原则，使用字符面积排序进行 block 组合
2. 对模型解决效果较差的样本（block 数目多于 4 个）进行冲采样
3. “尊重” OCR 识别文本结果，采用模糊匹配进行模型结果校正

#### 致谢

整个比赛过程中感谢认知智能实验室的同事以及实习生们给予的支持，感谢主办方精心准备赛题，让我们有机会能讲算法应用在有趣又具有挑战性的真实场景中。