

全球人工智能 2021 技术创新大赛 GLOBAL AI INNOVATION CONTEST

赛道一：医学影像报告异常检测

别再卷了



CONTENTS

1 团队背景和成员简介

2 赛题解读和数据分析

3 整体设计

4 创新和落地

5 方案总结



CONTENTS

1 团队背景和成员简介

2 赛题解读和数据分析

3 整体设计

4 创新和落地

5 方案总结



CONTENTS

1 团队背景和成员简介

2 赛题解读和数据分析

3 整体设计

4 创新和落地

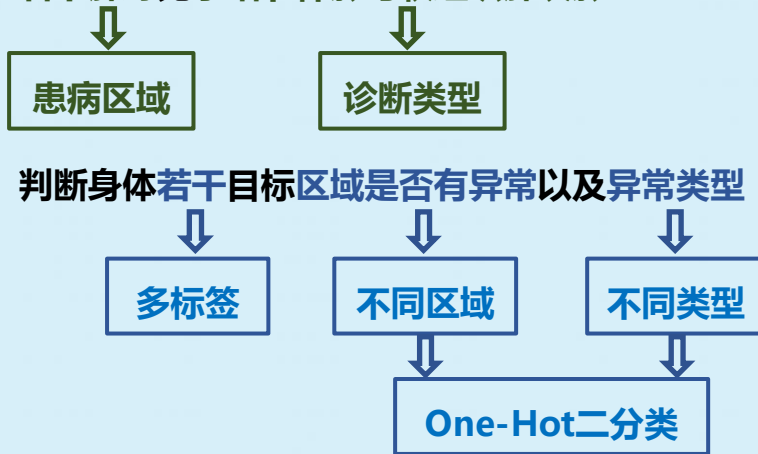
5 方案总结

赛题解读和数据分析

赛题解读

提供：CT影像描述**文本**

样例：右下肺野见小结节样影与软组织肿块影



需求：判断身体**若干目标区域**是否有异常以及**异常类型**

NLU 自然语言理解，多标签 && 二分类

2021 全球人工智能技术创新大赛

GLOBAL AI INNOVATION CONTEST

赛题解读和数据分析

赛题解读

赛题数据

列名	类型	示例
report_ID	int	1
description	文本, 脱敏	101 47 12 66 74 90 0 411 234 79
label	区域ID, 类型ID	3 4, 0 2

数据类型: 脱敏数据, 无明文 **if pretrain:**

Train from scratch!

评估方式: mlogloss

针对脱敏数据的分类问题

2021 全球人工智能技术创新大赛

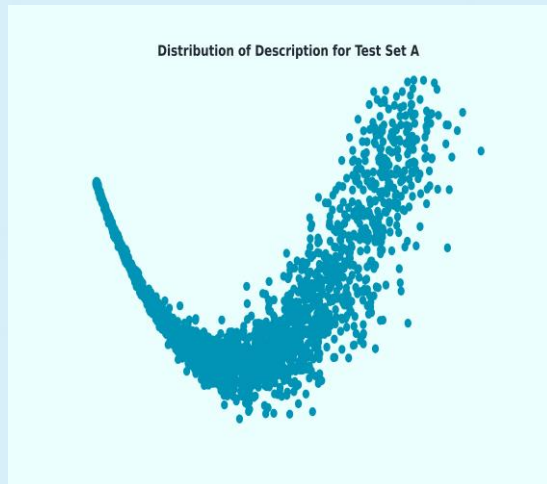
GLOBAL AI INNOVATION CONTEST

赛题解读和数据分析

数据分析



&



PCA降至二维可视化

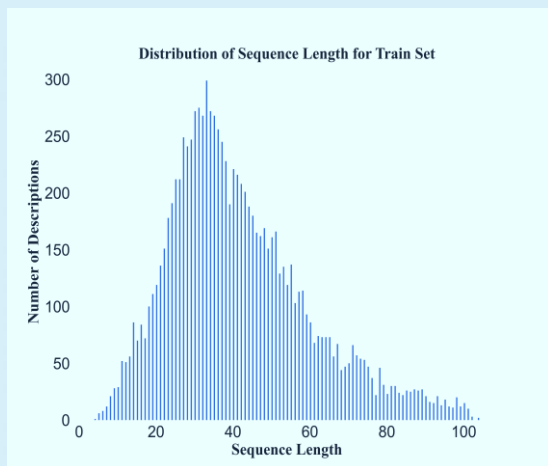
训练集和测试集同分布

2021全球人工智能技术创新大赛

GLOBAL AI INNOVATION CONTEST

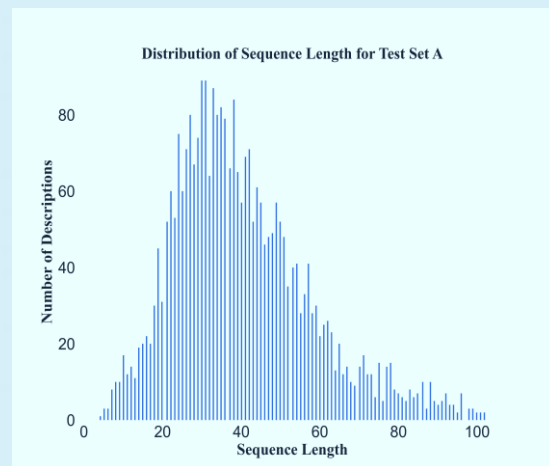
赛题解读和数据分析

数据分析



Min: 4 Max: 104 Mean: 41

&



Min: 4 Max: 102 Mean: 40



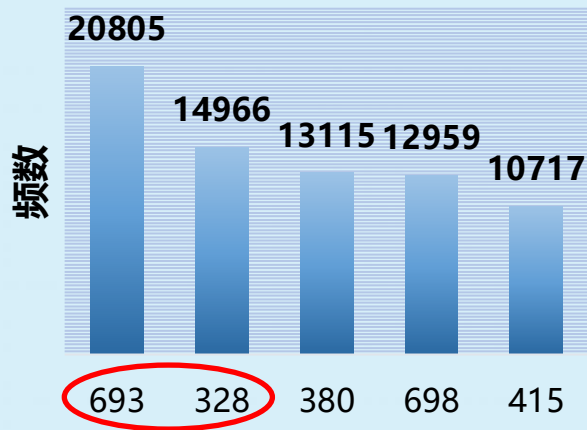
2021 全球人工智能技术创新大赛

GLOBAL AI INNOVATION CONTEST

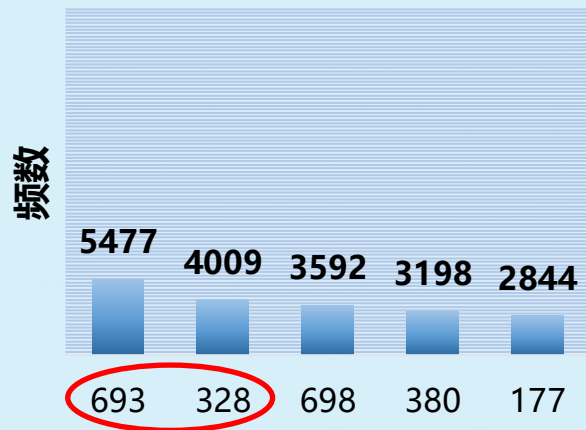
赛题解读和数据分析

数据分析

训练集 – 词频Top5



测试集A – 词频Top5



编码: 0~857 vocab_size=858



CONTENTS

1 团队背景和成员简介

2 赛题解读和数据分析

3 整体设计

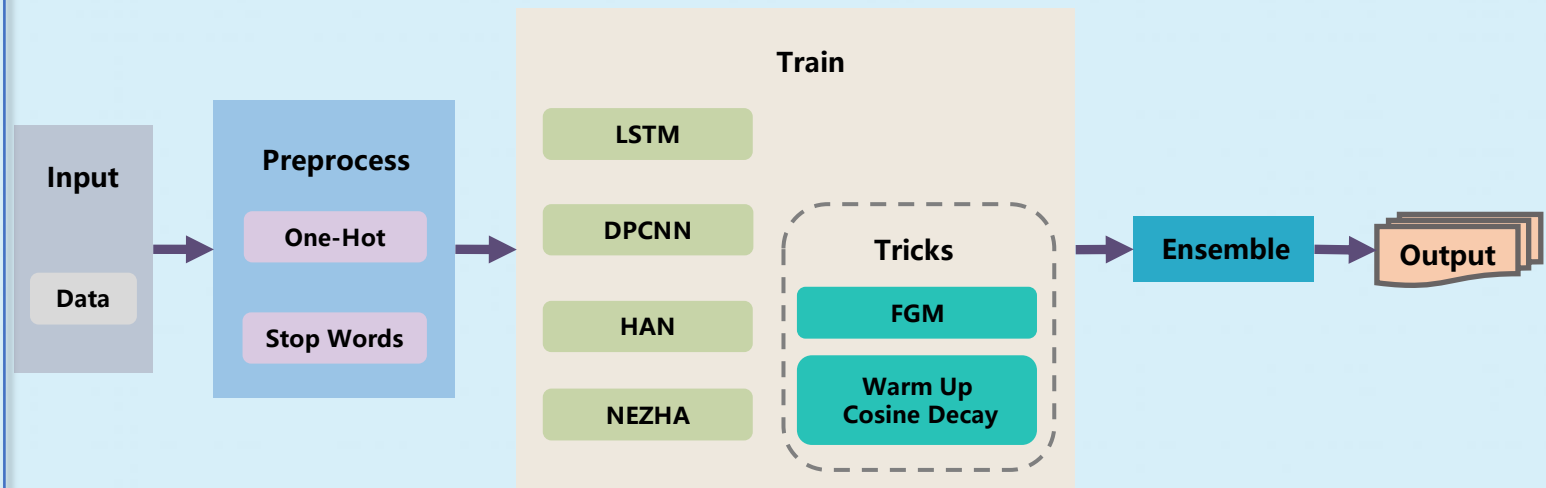
4 创新和落地

5 方案总结

整体设计

初赛-模型算法

Normandy: NEZHA, Long Short-Term Memory, Hierarchical Attention Networks and Deep Pyramid Convolutional Neural Networks

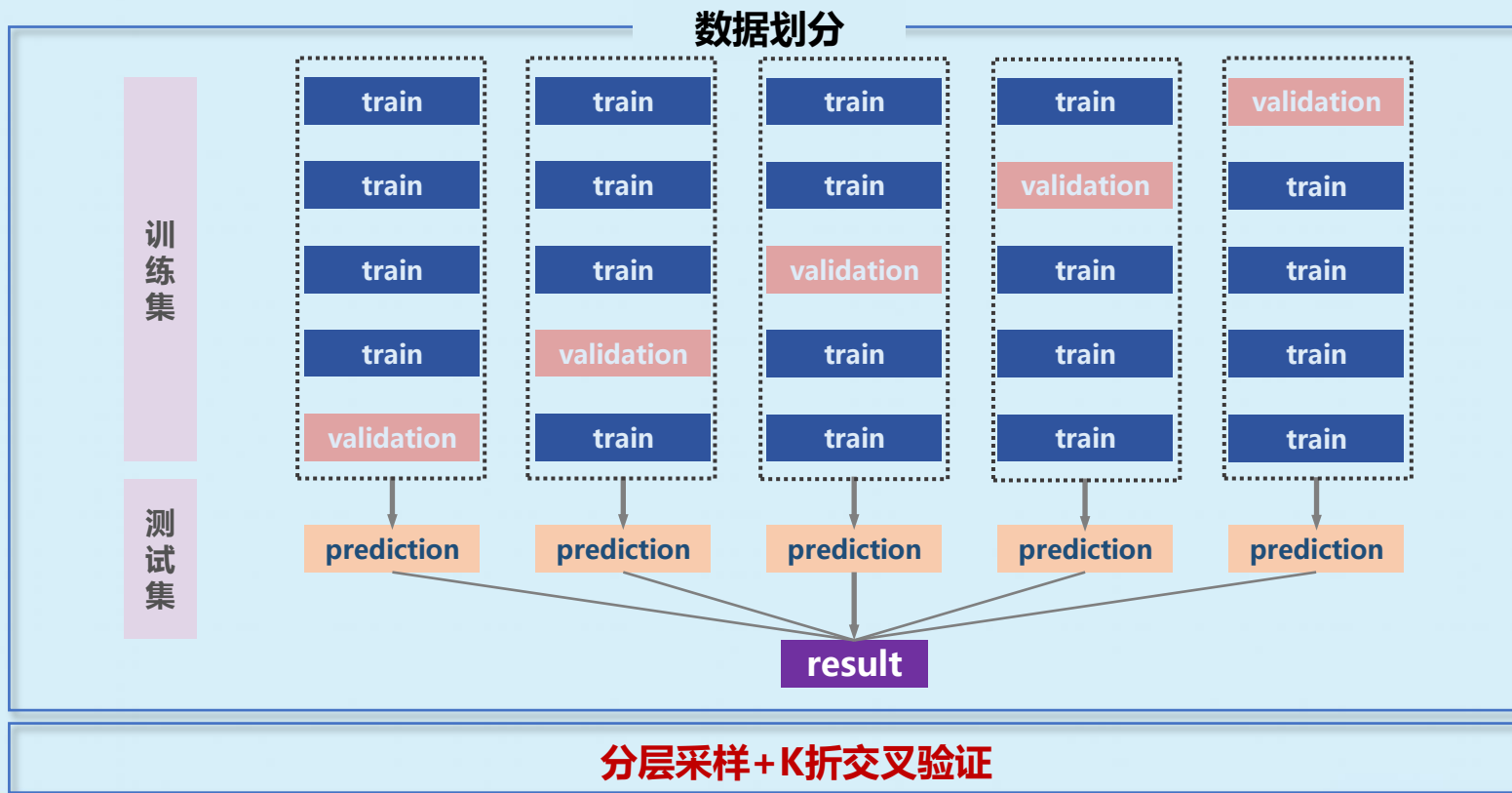


诺曼底

2021全球人工智能技术创新大赛

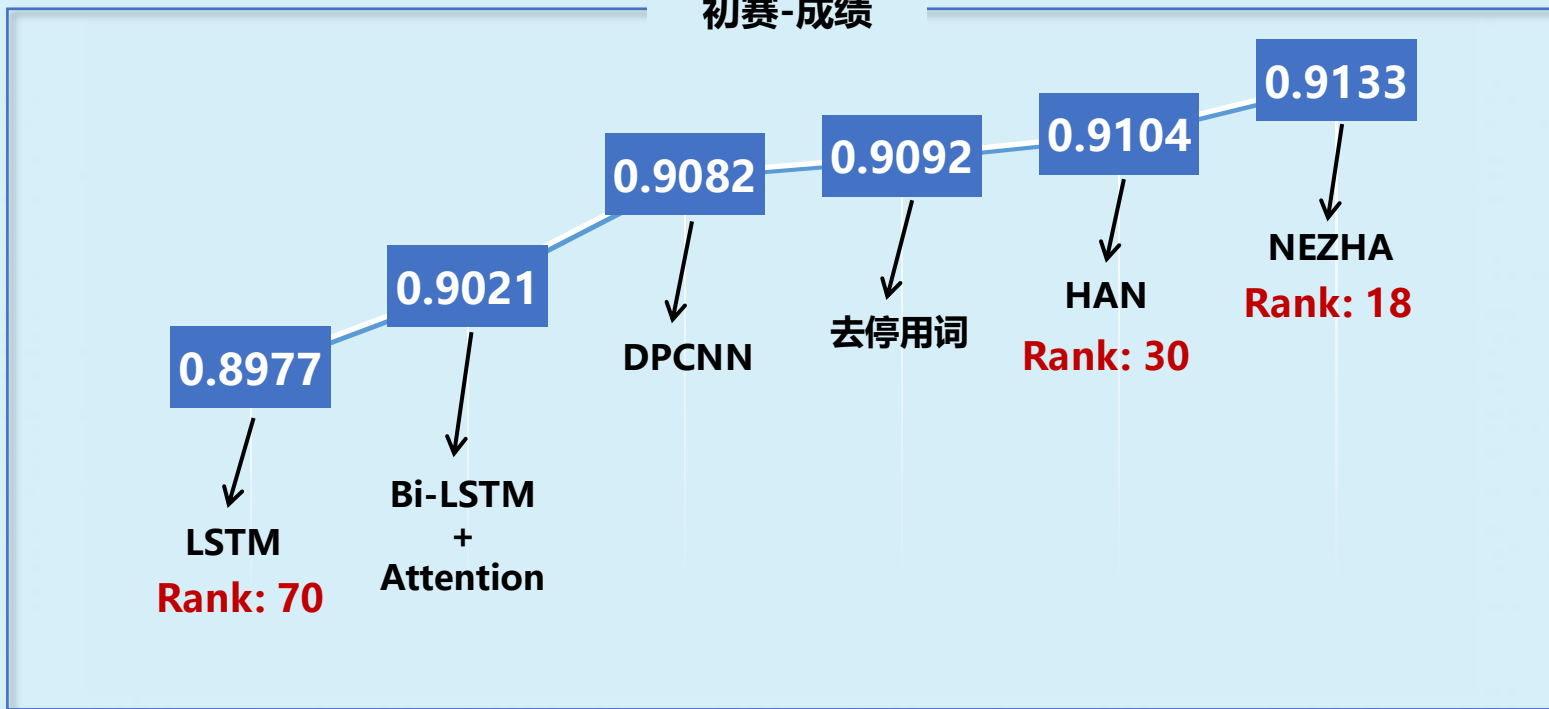
GLOBAL AI INNOVATION CONTEST

整体设计



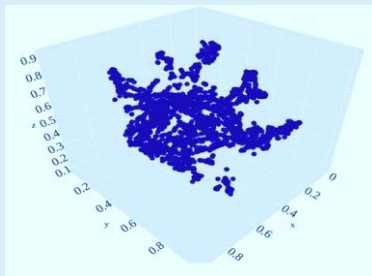
整体设计

初赛-成绩

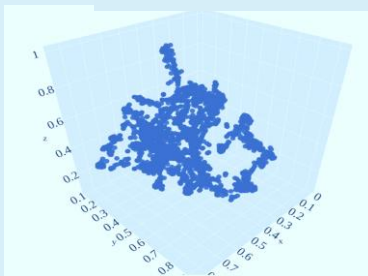


整体设计

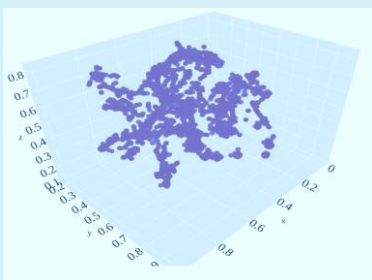
结果可视化



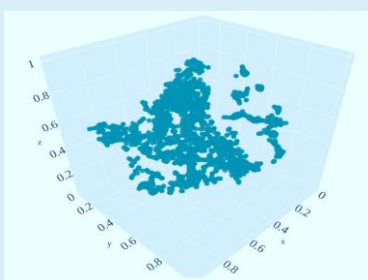
DPCNN



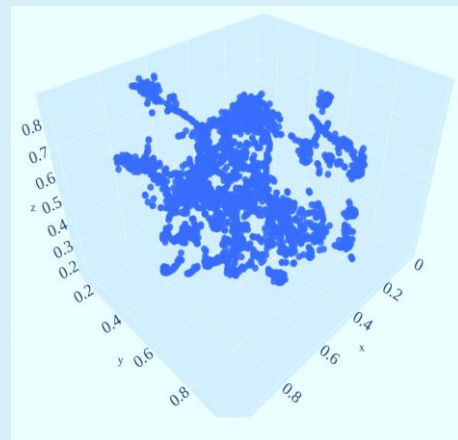
LSTM



HAN



NEZHA



merge



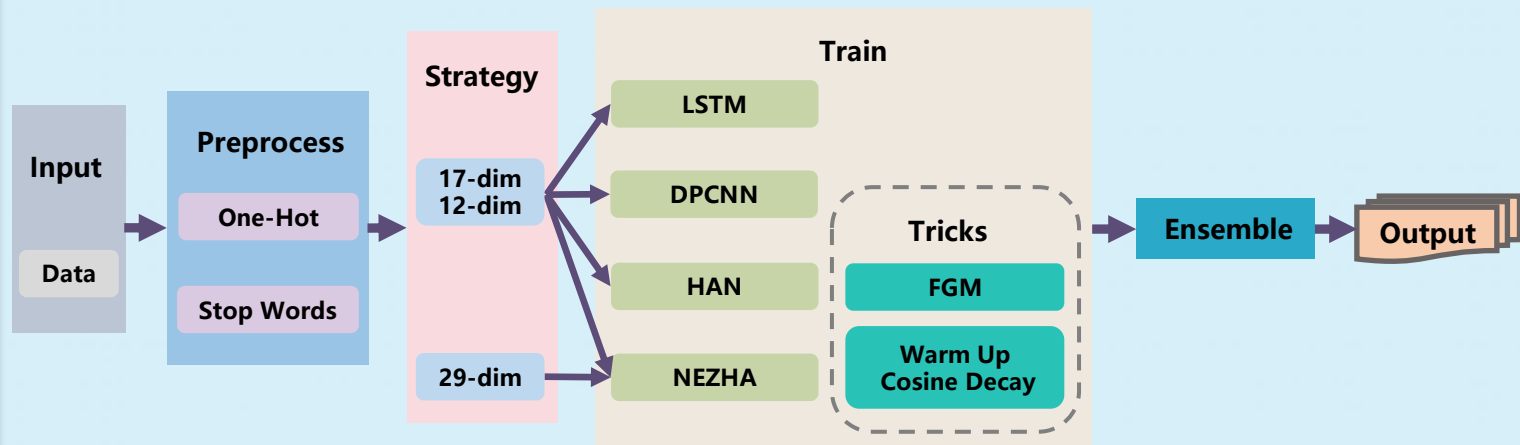
2021 全球人工智能技术创新大赛

GLOBAL AI INNOVATION CONTEST

整体设计

复赛-模型算法

Noah's Ark: NEZHA Trained Jointly and Separately with HAN, LSTM and Deep Pyramid Convolutional Neural Networks



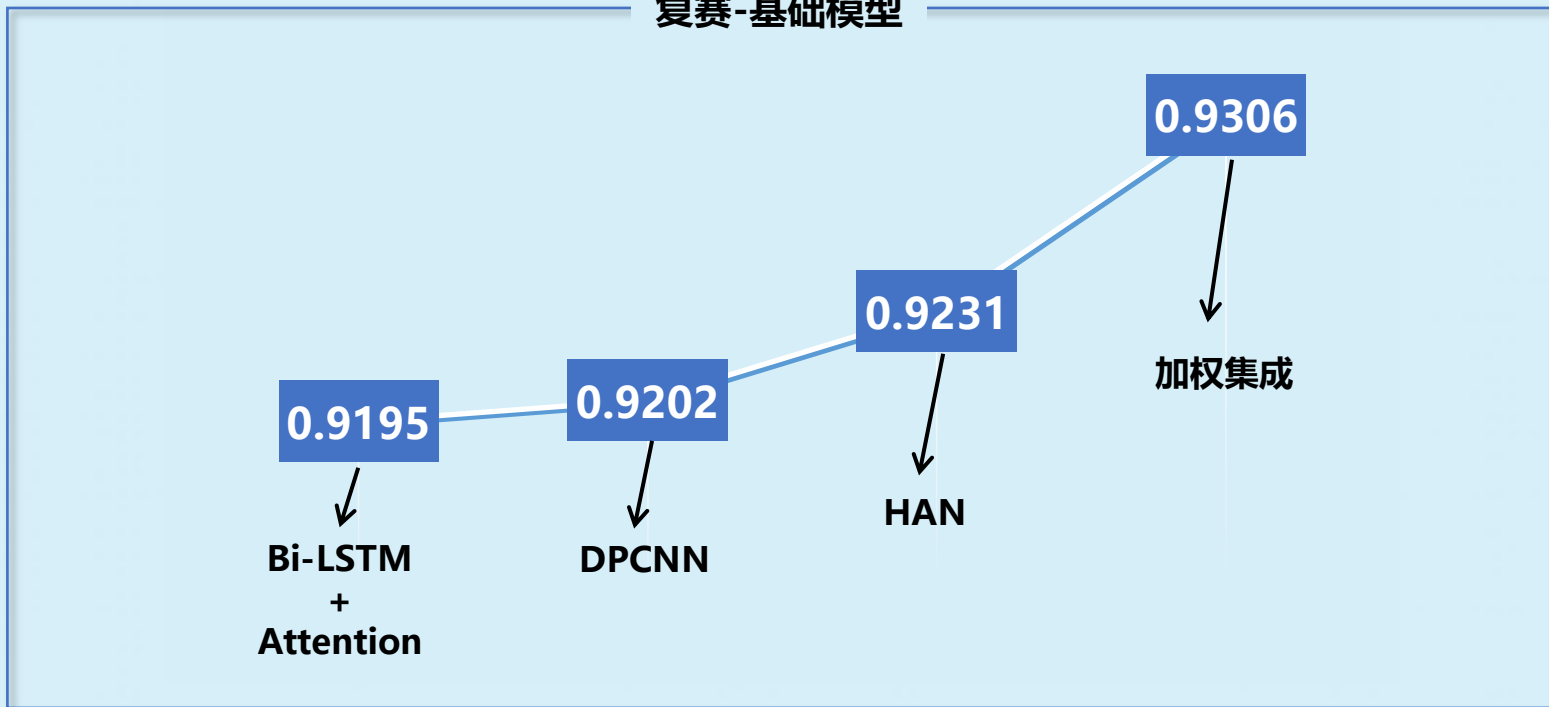
诺亚方舟

2021 全球人工智能技术创新大赛

GLOBAL AI INNOVATION CONTEST

整体设计

复赛-基础模型

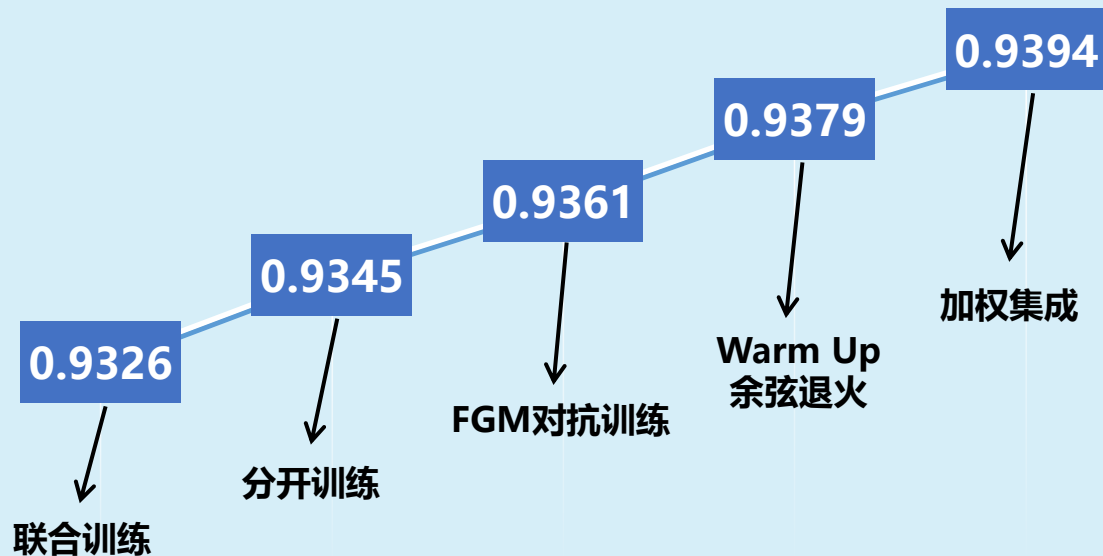


2021 全球人工智能技术创新大赛

GLOBAL AI INNOVATION CONTEST

整体设计

复赛-NEZHA



2021 全球人工智能技术创新大赛

GLOBAL AI INNOVATION CONTEST

整体设计

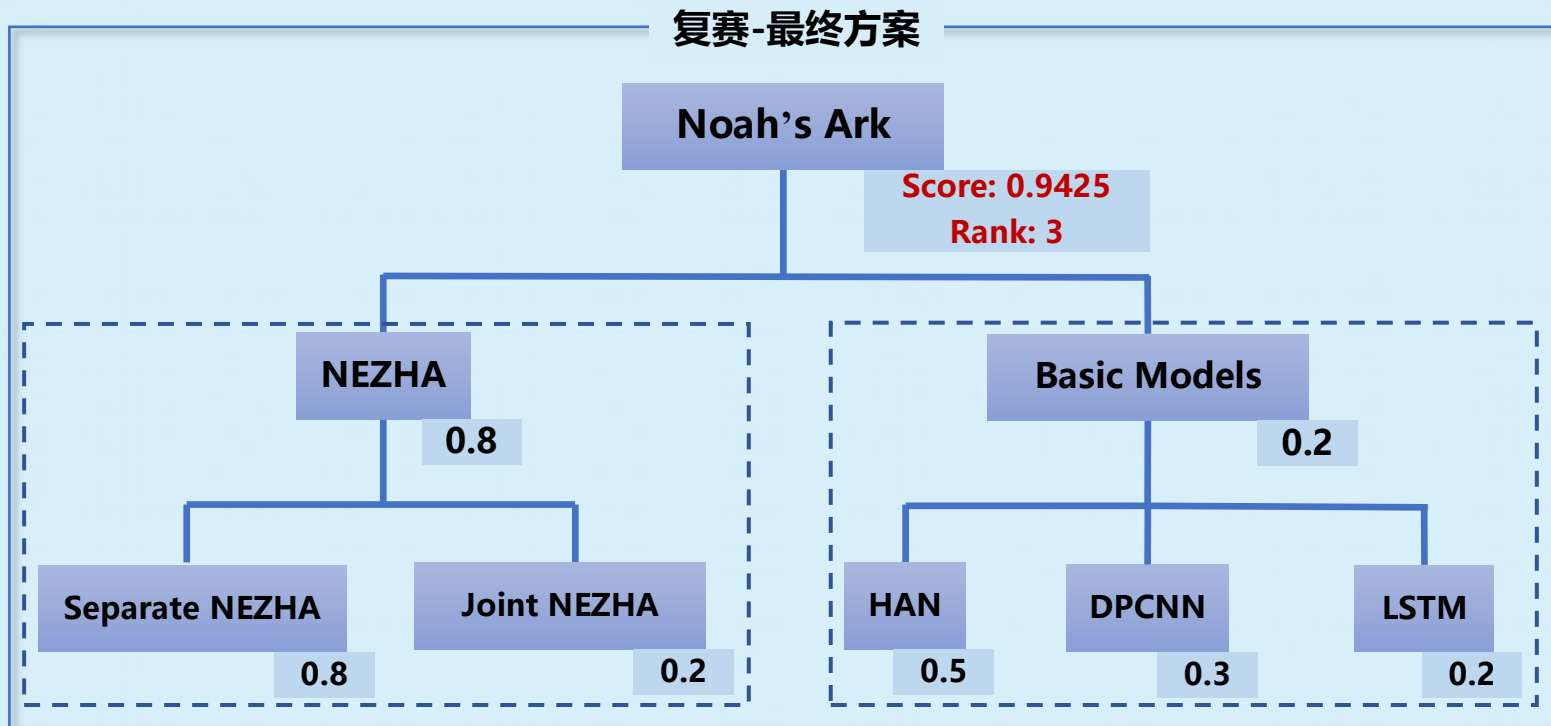
NEZHA训练策略

模型	策略	任务1得分	任务2得分	总得分	任务gap
NEZHA	联合训练	0.9257	0.9142	0.9212	0.0115
NEZHA	分开训练	0.9326	0.9302	0.9316	0.0024

使用分开训练总得分上升，两个任务的得分差距缩小



整体设计





CONTENTS

1 团队背景和成员简介

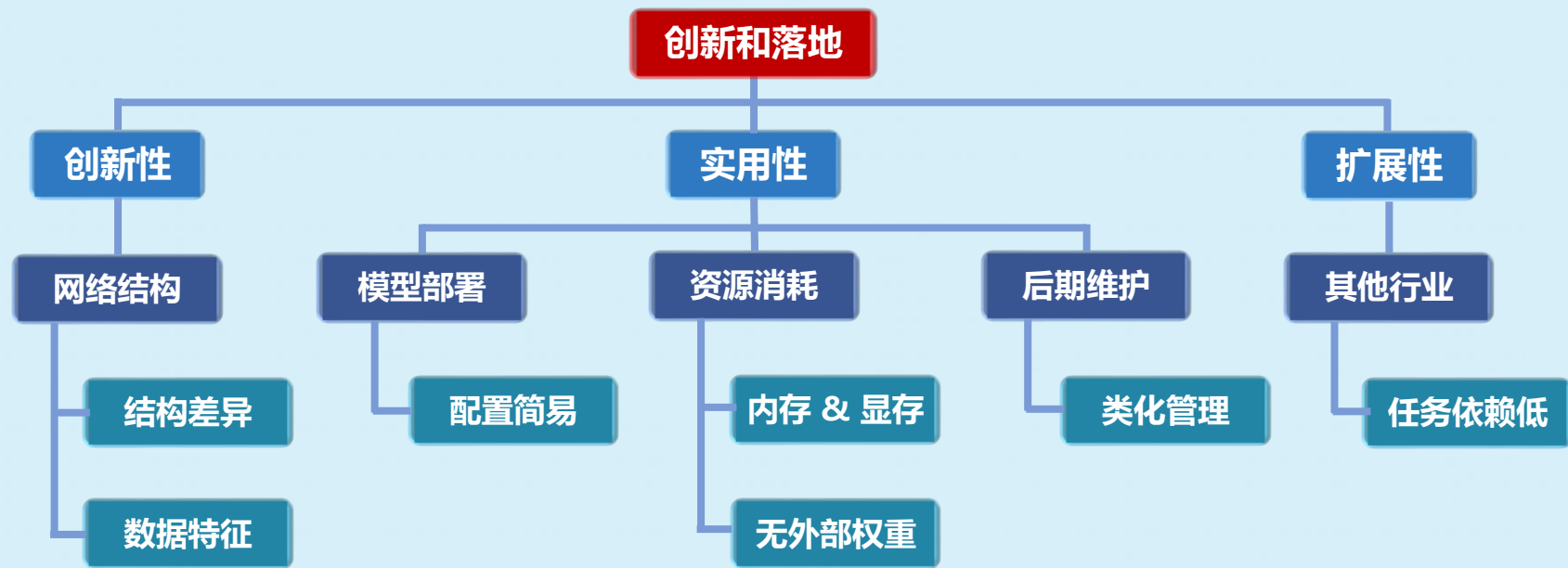
2 赛题解读和数据分析

3 整体设计

4 创新和落地

5 方案总结

创新和落地





CONTENTS

1 团队背景和成员简介

2 赛题解读和数据分析

3 整体设计

4 创新和落地

5 方案总结

方案总结

管理

参数管理

版本管理

分工合作

模型

匹配

多样性

尝试



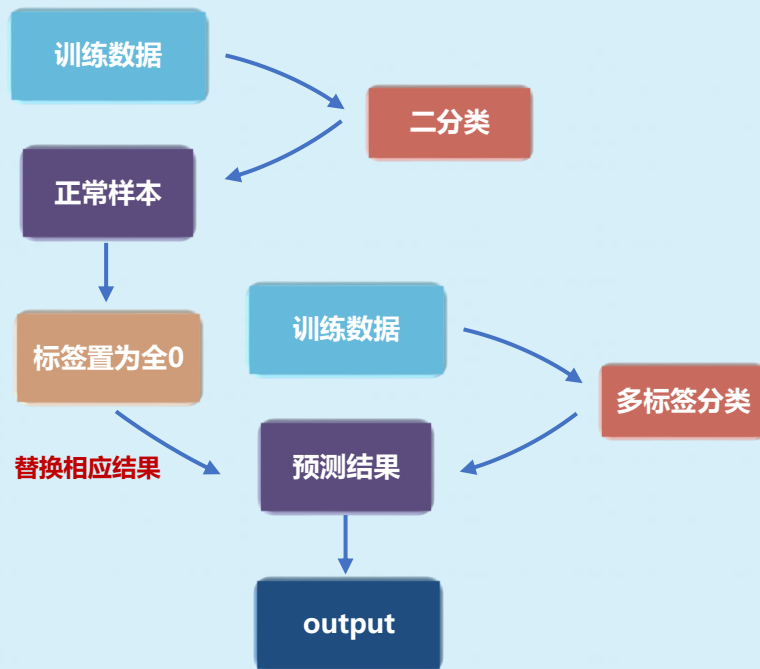
2021 全球人工智能技术创新大赛

GLOBAL AI INNOVATION CONTEST

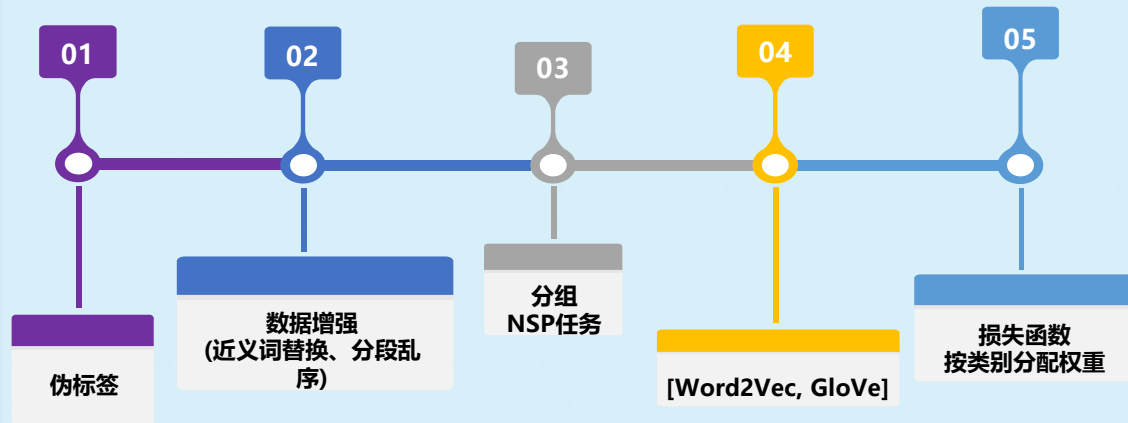
方案总结



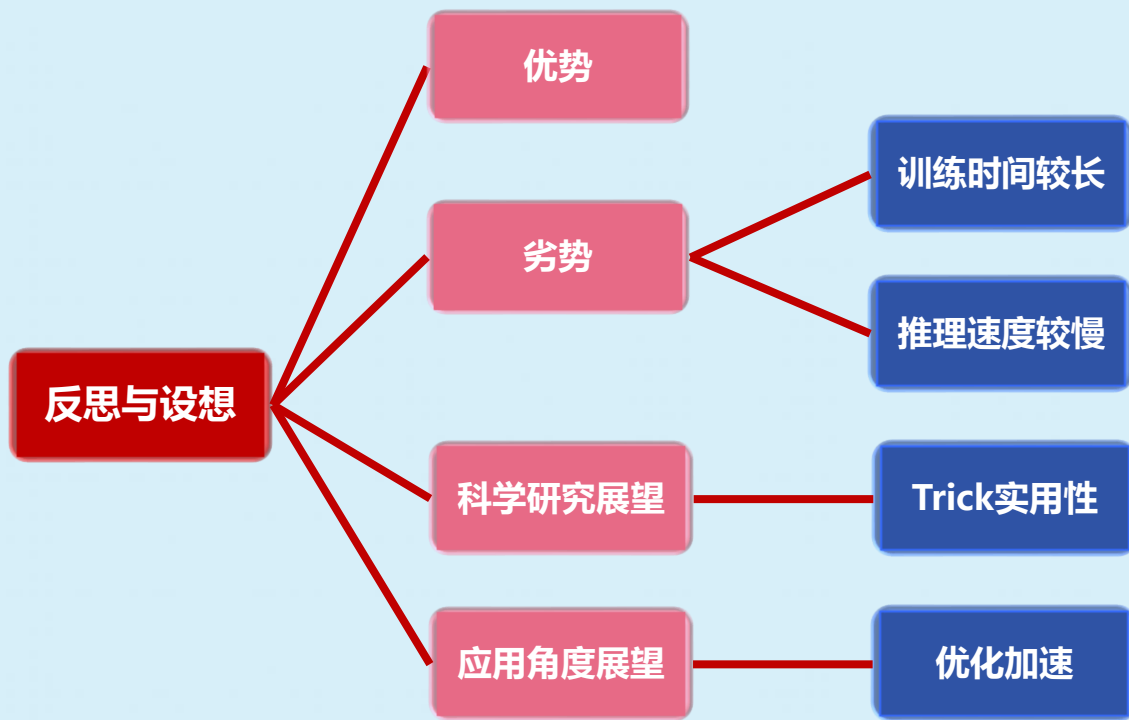
Two to Multi



方案总结



方案总结





THANKS!

附录

数据分析

各类标签出现次数



类别不平衡，重加权(re-weighting)