

WPS AI

WPS AI

WPS AI

文件

开始

插入

页面

引用

审阅

视图

工具

会员专享

WPS AI

白格式刷

字体(正文)

大小

加粗

斜体

下划线

删除线

颜色

背景色

边框

段落

列表

查找

替换

智能公文

一、神经网络 (24%)

神经网络 (Neural Network) 是一种机器学习程序或模型, 它以类似于人脑的方式做出决策, 通过使用模仿生物神经元协同工作方式的过程来识别现象、权衡利弊并得出结论。其完整的应用流程通常包括模型训练、推理、评估等过程。

根据以上材料, 回答下列问题 (非学习同学完成 1-4 题; 学习同学完成 3-8 题):

- 训练一个神经网络模型时, 需要确定损失函数, 其中交叉熵损失 (Cross-Entropy Loss) 是分类任务中常用的损失函数。假设在二分类场景中, 正确答案为 y , 模型预测值为 $\hat{y}$, 请直接给出此时交叉熵损失的计算公式。(2%)
- 训练神经网络模型时, 可以通过建立计算图进行高效的数值微分操作。现给定以下计算图, 请求解 L , $\frac{\partial L}{\partial a}$, $\frac{\partial L}{\partial b}$, $\frac{\partial L}{\partial c}$, 其中 $\sigma(x) = \frac{1}{1+e^{-x}}$ 。注意保留所有中间计算过程。(8%)

$a = e$

$h = \ln a$

$b = 9$

$m = b + h$

WPS AI

WPS AI

WPS AI

文件

开始

插入

页面

引用

审阅

视图

工具

会员专享

WPS AI

白格式刷

字体(正文)

大小

加粗

斜体

下划线

删除线

颜色

背景色

边框

段落

列表

查找

替换

智能公文

二、神经网络模型训练 (24%)

神经网络 (Neural Network) 是一种机器学习程序或模型, 它以类似于人脑的方式做出决策, 通过使用模仿生物神经元协同工作方式的过程来识别现象、权衡利弊并得出结论。其完整的应用流程通常包括模型训练、推理、评估等过程。

根据以上材料, 回答下列问题 (非学习同学完成 1-4 题; 学习同学完成 3-8 题):

- 训练一个神经网络模型时, 需要确定损失函数, 其中交叉熵损失 (Cross-Entropy Loss) 是分类任务中常用的损失函数。假设在二分类场景中, 正确答案为 y , 模型预测值为 $\hat{y}$, 请直接给出此时交叉熵损失的计算公式。(2%)
- 训练神经网络模型时, 可以通过建立计算图进行高效的数值微分操作。现给定以下计算图, 请求解 L , $\frac{\partial L}{\partial a}$, $\frac{\partial L}{\partial b}$, $\frac{\partial L}{\partial c}$, 其中 $\sigma(x) = \frac{1}{1+e^{-x}}$ 。注意保留所有中间计算过程。(8%)
- 束搜索解码 (Beam Search Decoding) 算法是一种在模型推理中常用的解码策略, 假设现在有如下的转移对数似然表, 其中 $\langle \text{Start} \rangle$ 和 $\langle \text{End} \rangle$ 分别为开始和结束标志, 请在 BeamSize 为 2 的条件下根据该表进行解码, 注意给出每一步解码时保留的候选序列, 并给出最终输出的解码序列。(8%)

$a = e$

$h = \ln a$

$b = 9$

$m = b + h$

$c = 0$

$p = \sigma(c)$

$L = m \cdot p$

$x \rightarrow y$	Natural	Science	Language	Text	Processing	Addressing	$\langle \text{End} \rangle$
$\langle \text{Start} \rangle \rightarrow$	0.5	0.3	0.2	0.1	0.4	0.6	0.0
$\text{Natural} \rightarrow$	0.8	0.4	0.3	0.2	0.5	0.7	0.0
$\text{Science} \rightarrow$	0.6	0.7	0.4	0.3	0.6	0.8	0.0
$\text{Language} \rightarrow$	0.4	0.5	0.6	0.5	0.7	0.9	0.0
$\text{Text} \rightarrow$	0.3	0.6	0.5	0.7	0.8	1.0	0.0
$\text{Processing} \rightarrow$	0.2	0.5	0.4	0.6	0.9	1.0	0.0
$\text{Addressing} \rightarrow$	0.1	0.4	0.3	0.5	0.8	1.0	0.0
$\langle \text{End} \rangle \rightarrow$	0.0	0.3	0.2	0.1	0.4	0.6	0.0

$c = 0$ $p = \sigma(c)$

3. 束搜索解码 (Beam Search Decoding) 算法是一种在模型推理中常用的解码策略, 假设现在有如下的转移对数似然表, 其中<Start>和<End>分别为开始和结束标志, 请在 Beamsize 为 2 的条件下根据该表进行解码, 注意给出每一步解码时保留的候选序列, 并给出最终输出的解码序列。(8%)

x→y	Natural	Science	Language	Text	Processing	Addressing	<End>
<Start>	-0.5	-0.9	-1.7	-1.5	-2.2	-1.3	-3
Natural			-1.2	-1.8			
Science			-1.6	-1.3			
Language					-0.9	-1.3	
Text					-0.4	-2.3	
Processing							-0.5
Addressing							-0.4

4. BLEU 得分是评估机器翻译质量的常用指标, 其具体计算过程如下:

- 1) 计算机器译文中与参考译文共有的 n -gram 的精确度 (即比例) p_n ;
- 2) 计算不同 n -gram 精确度的几何平均值 $p = \exp(\sum_{n=1}^N \frac{1}{N} \log(p_n))$;

Addressing

4. BLEU 得分是评估机器翻译质量的常用指标, 其具体计算过程如下:

- 1) 计算机器译文中与参考译文共有的 n -gram 的精确度 (即比例) p_n ;
- 2) 计算不同 n -gram 精确度的几何平均值 $p = \exp(\sum_{n=1}^N \frac{1}{N} \log(p_n))$;
- 3) 定义短句惩罚 $BP = 1$ (if $c > r$ else $e^{(1-r/c)}$), 其中 c 是机器译文的长度, r 是最接近 c 的参考译文的长度, 则最终得分 $BLEU = BP \cdot p$.

现有如下机器译文及其参考译文:

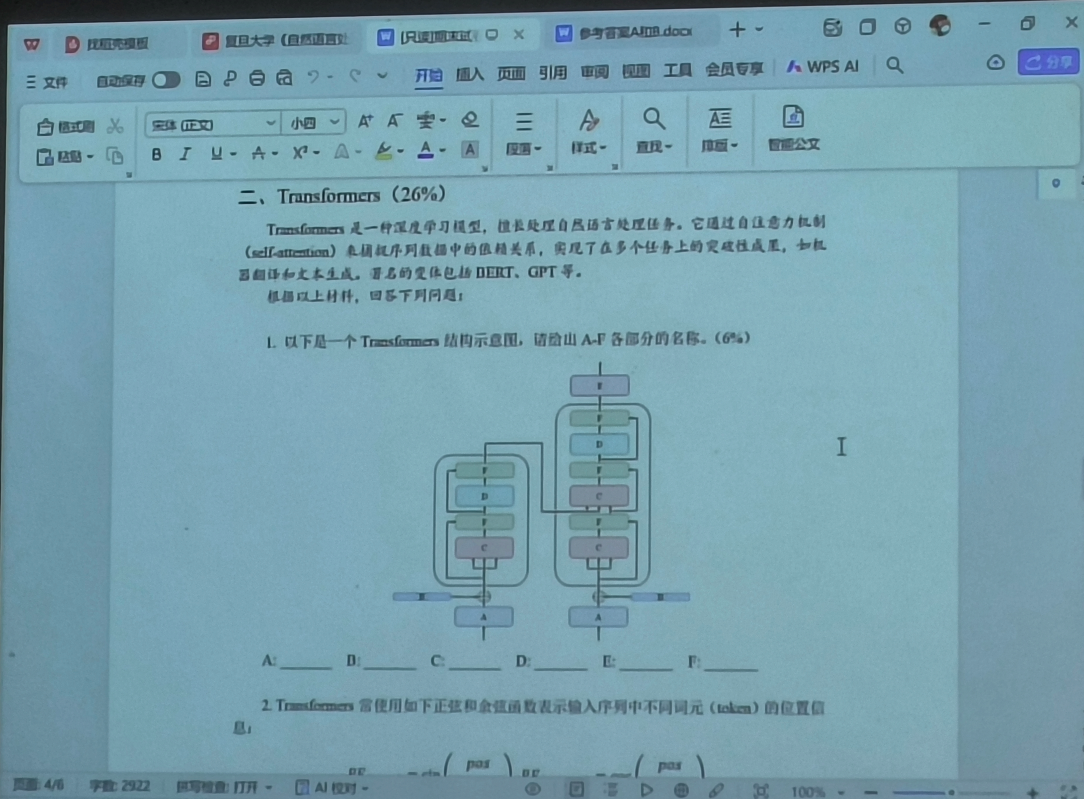
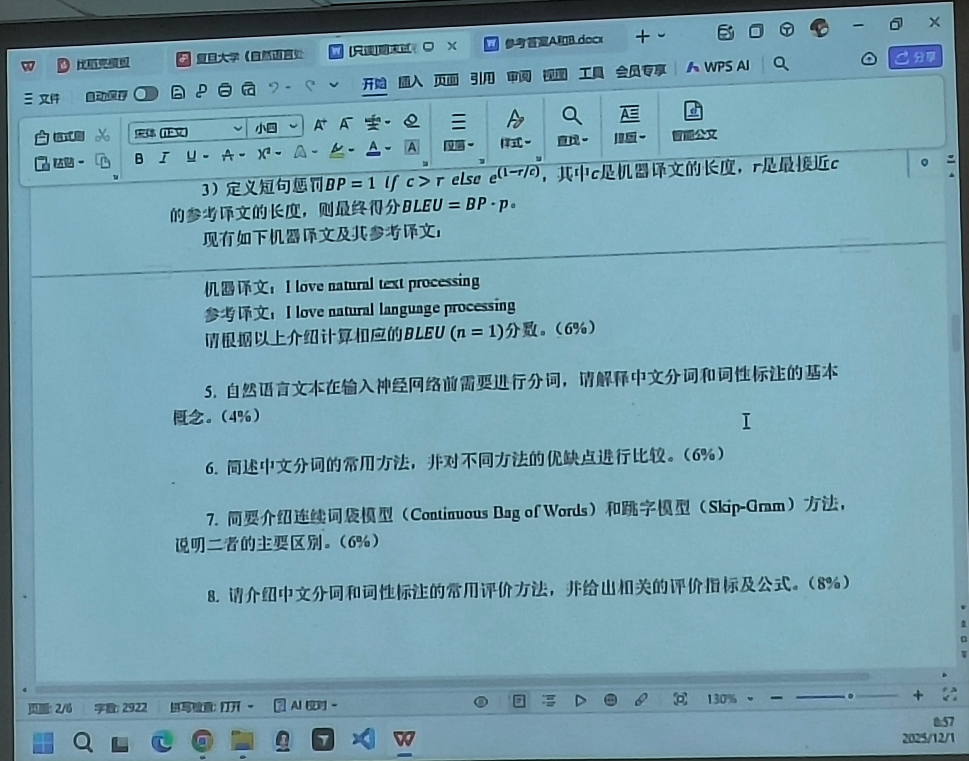
机器译文: I love natural text processing
 参考译文: I love natural language processing

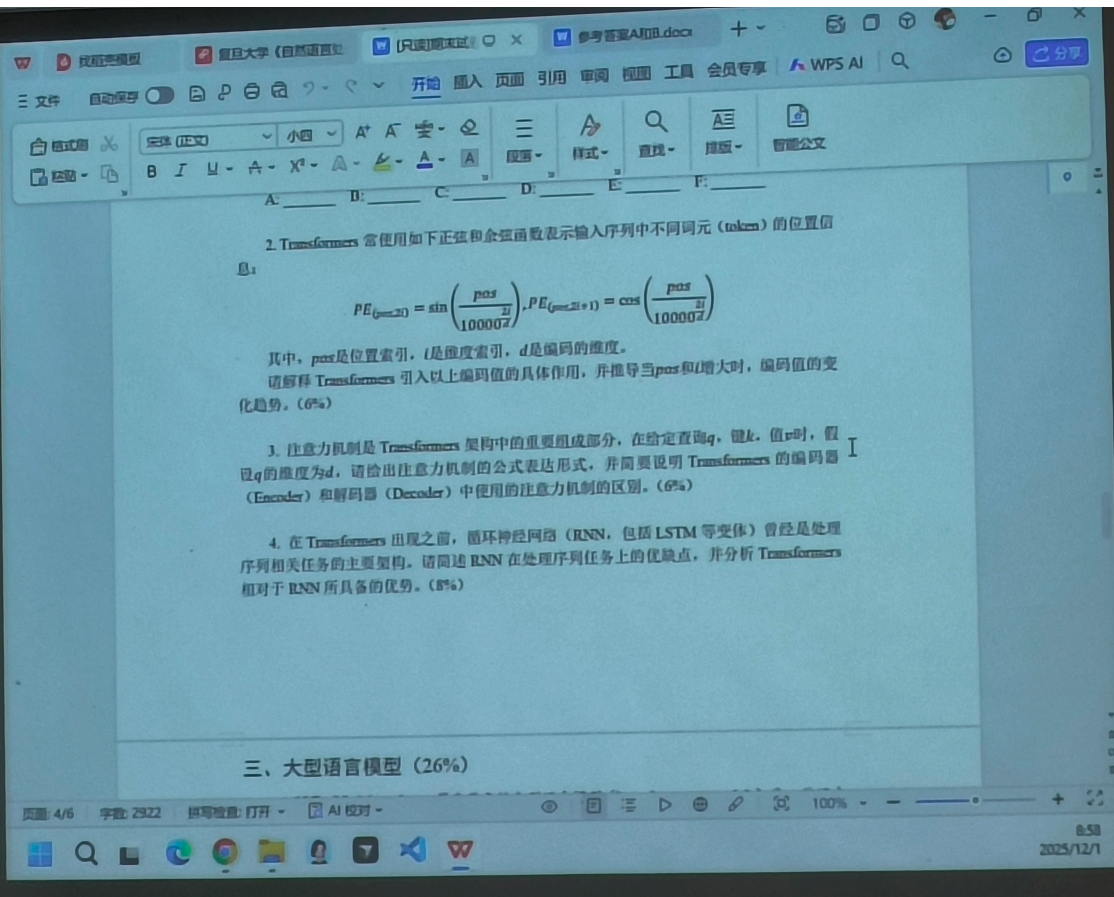
请根据以上介绍计算相应的 BLEU ($n = 1$) 分数。(6%)

5. 自然语言文本在输入神经网络前需要进行分词, 请解释中文分词和词性标注的基本概念。(4%)

6. 简述中文分词的常用方法, 并对不同方法的优缺点进行比较。(6%)

7. 简要介绍连续词袋模型 (Continuous Bag of Words) 和跳字模型 (Skip-Gram) 方法,





(3) 区别: CROW 模型通过上下文输入, 输入目标词的向量, 输出是目标词; 而 Skip-gram 模型通过目标词来预测上下文词, 输入是目标词, 输出是上下文词。类似表达即可, 表达不清扣 1 分, 错误得 0 分。
2 分, 类似表达即可, 表达不清扣 1 分, 错误得 0 分。

找稻壳模板

复旦大学《自然语言处理和知识表示》

[只读]期末试卷.docx

开始 插入 页面 引用 审阅

A 卷参考答案

一、神经网络 (24%)

非学程同学完成 1-4 题; 学程同学完成 5-8 题

1. $L = -y \log \hat{y} - (1-y) \log(1-\hat{y})$

书写正确得 2 分, 忘记取负得 1 分, 错误得 0 分

2. (1) $h = \ln a = \ln e = 1$

0.5 分

(2) $m = b + h = 9 + 1 = 10$

0.5 分

(3) $p = \sigma(c) = \frac{1}{1+e^0} = \frac{1}{2}$

0.5 分

(4) $L = m * p = 10 * \frac{1}{2} = 5$

1 分

(5) $\frac{\partial L}{\partial m} = p = \frac{1}{2}$

0.5 分

(6) $\frac{\partial L}{\partial p} = m = 10$

0.5 分

(7) $\frac{\partial L}{\partial h} = \frac{\partial L}{\partial m} \cdot \frac{\partial m}{\partial h} = \frac{1}{2} \cdot 1 = \frac{1}{2}$

0.5 分

(8) $\frac{\partial L}{\partial a} = \frac{\partial L}{\partial h} \cdot \frac{\partial h}{\partial a} = \frac{1}{2} \cdot \frac{1}{e} = \frac{1}{2e}$

1 分

(9) $\frac{\partial L}{\partial b} = \frac{\partial L}{\partial m} \cdot \frac{\partial m}{\partial b} = \frac{1}{2} \cdot 1 = \frac{1}{2}$

1 分

(10) $\frac{\partial L}{\partial c} = \frac{\partial L}{\partial p} \cdot \frac{\partial p}{\partial c} = 10 \cdot (p \cdot (1-p)) = 10 \cdot (\frac{1}{2} \cdot \frac{1}{2}) = \frac{5}{2}$

1 分

整体过程保留完整 1 分

3. (1) <Start> Natural -0.5 (保留)

<Start> Science -0.9 (保留)

<Start> Language -1.7

<Start> Text -1.5

<Start> Processing -2.2

<Start> Addressing -1.3

<Start> <End> -3

(2) <Start> Natural Language -0.5-1.2=-1.7 (保留)

<Start> Natural Text -0.5-1.8=-2.3

<Start> Science Language -0.9-1.6=-2.5

<Start> Science Text -0.9-1.3=-2.2 (保留)

(3) <Start> Natural Language Processing -1.7-0.9=-2.6 (保留)

<Start> Natural Language Addressing -1.7-1.3=-3

<Start> Science Text Processing -2.2-0.4=-2.6 (保留)

<Start> Science Text Addressing -2.5-2.3=-4.8

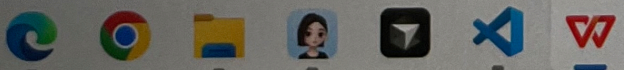
(4) <Start> Natural Language Processing <End> -2.6-0.5=-3.1 (最终解码序列)

<Start> Science Text Processing <End> -2.6-0.5=-3.1

每步 2 分, 给出完整解码过程和保留序列 2 分, 给出部分解码过程给 1 分, 解码错误或保留序列错误 0 分。

4. (1) $p_1 = 4/5$ (2) $p = \exp(\log p_1) = \exp(\log 4/5) = 4/5$

7 拼写检查: 打开 AI 校对 缺失字体



(3) $BP = 1, BLUE = BP \cdot p = \exp(\log 4/5) = 4/5$

每步 2 分, 计算正确得 2 分, 错误得 0 分。

5. (1) 中文分词: 中文分词是指将连续字序列转换为对应的词序列的过程, 也可以看做在输入的序列中添加空格或其他边界标记的过程。 2 分, 类似表达即可, 表达不清晰得 1 分, 错误得 0 分。

(2) 词性标注: 词性标注是指在给定的语境中确定句子中各词的词性。 2 分, 类似表达即可, 表达不清晰得 1 分, 错误得 0 分。

6. (1) 基于最大匹配的中文分词方法: (优点) 简单、快速、可控、仅依赖词表; (缺点) 对于没有在词典中出现的词没有很好的处理方案, 同时对于分词歧义的处理能力也不足。

(2) 基于线性链条件随机场的中文分词: (优点) 条件随机场 (CRF) 能够建模序列标签之间的依赖关系, 较好地处理上下文信息; (缺点) 训练代价大、复杂度高。

(3) 基于感知器的中文分词方法: (优点) 简单高效; (缺点) 准确性有限。

(4) 基于双向长短期记忆网络的中文分词方法: (优点) 有效捕捉长距离依赖关系; (缺点) 需要较大的计算资源进行训练。

6 分, 至少回答 2 种方法, 并给出优缺点。每种方法名称 1 分, 优点 1 分, 缺点 1 分。

7. (1) 连续词袋模型假设文本中的词可以通过其在文本中的上下文词推导出来。CBOW 模型通过最大化中心词在上下文中出现的条件概率, 来进行模型的训练。 2 分, 类似表达即可, 表达不清晰得 1 分, 错误得 0 分。

(2) 跳字模型的基本假设是文本中的词可以预测其上下文窗口内的其他词。通过最大化上下文词在给定中心词时的条件概率, 来进行模型的训练。 2 分, 类似表达即可, 表达不清晰得 1 分, 错误得 0 分。

(3) 区别: CBOW 模型通过上下文词来预测目标词, 输入是上下文词的向量, 输出是目标词; 而 Skip-Gram 模型通过目标词来预测上下文词, 输入目标词的向量, 输出是上下文词。 2 分, 类似表达即可, 表达不清晰得 1 分, 错误得 0 分。

8. (1) 中文分词: 精确率 (Precision, P)、召回率 (Recall, R)、F 值 (F-Measure)。

$$\text{精确率 (P)} = \frac{\text{算法输出的正确分词结果个数}}{\text{算法输出的全部分词结果个数}} \times 100\%$$

$$\text{召回率 (R)} = \frac{\text{算法输出的正确分词结果个数}}{\text{测试集中全部答案个数}} \times 100\%$$

$$F_\beta = \frac{(\beta^2 + 1) \times P \times R}{P + R} \times 100\%$$

(2) 词性标注: 准确率 (Accuracy) 和宏平均 F1 (Macro-F1)。

$$\text{准确率 (Accuracy)} = \frac{\text{算法输出的正确结果个数}}{\text{算法输出的全部结果总数}} \times 100\%$$

$$P_{\text{名词}} = \frac{\text{算法输出的正确名词结果个数}}{\text{算法输出的名词结果总数}} \times 100\%$$

$$R_{\text{名词}} = \frac{\text{算法输出的正确名词结果个数}}{\text{测试集中正确名词结果总数}} \times 100\%$$

电池状态: 已充电
节能模式开启



$$Macro-F1 = \frac{1}{n} \sum_{i=1}^n F1_i$$

每项为4分,其中,给出评价标准得1分,给出相应公人得3分,不完整适当扣分。

二、Transformers (26%)

1. A. 词嵌入 B. 位置编码 C. 多头注意力 D. 前馈层 E. 残差连接 F. 线性层

每句1分。若原文中文超过翻译也正确。错误得0分。

2. (1)作用: 确保在不同难度上位置编码具有不同的周期性变化, 捕捉序列中词语的位置信息。2分, 类似表达即可, 表达不准确适当扣分。未回答到位置信息得0分。

(2) 变化趋势: 位置变化 pos: 编码值周期性变化; 准度变化 $\pm$: 编码值变化频率变
小, 不同值度上周期不同。4 分, 有假乎过程并得到类似表达即可, 没有假乎过程最多得
2 分。

3. (1) $A_{\text{eff}} = \frac{e k^2}{\sqrt{V}} v$ 2分, 适当变形不扣分, 没有由一化得1分, 写错得0分.

(2) 编解码：多头自注意力；解码器：掩码多头自注意力，交叉多头注意力 4 分，每由三种注意力为正确项得分满分。少答一种扣 1 分，多答不得分。

4. (1) 优点: 捕捉序列信息; 长短期记忆 (LSTM) 和门控循环单元 (GRU)

(2) 缺点: 序列依赖; 长期依赖问题; 梯度问题

(3) 优势：并行处理；长程依赖；位置编码；梯度问题；扩展性和灵活性；扩展性
8 分。答出一个要点得 1 分，每个小部分最多得 3 分。

三、大型语言模型 (26%)

1. (1) 学习过程: 预训练, SFT, RLHF 6分, 每答出一个阶段得2分。

(2) 类人回复: 对 RLHF 的介绍 2 分。根据表述清晰程度, 酌情不给分。

2. (1) 设计原则: 6分, 至少给出三点, 言之有理即可。缺少一点扣2分。

(2) 提示词优化: 2 分, 结合所给出的设计原则即可。

3. (1) 主要用途: 6分, 至少给出3点, 言之有理即可, 缺少一点扣2分。

(2) 不足之处: 2 分, 至少给出 1 点, 言之有理即可。

四、前沿進展 (24%)

非学期同学 2-4 中选择两道完成; 学期同学 1-4 中选择两道完成。

1.12 分，言之有理即可，没有具体案例最多给 6 分。

2.12分。言之有理即可。没有具体案例最多给6分。

3.12分。言之有理即可，至少谈3点，缺少一点扣4分。

4. (1) 作用: 4 分, 言之有理即可。

(2) 发展形态: 4分, 言之有理即可。

(3) 应用模式: 4分, 言之有理即可。

