



(12) 发明专利

(10) 授权公告号 CN 113378552 B

(45) 授权公告日 2024. 04. 19

(21) 申请号 202110760314.1

G06T 9/00 (2006.01)

(22) 申请日 2021.07.06

(56) 对比文件

(65) 同一申请的已公布的文献号

申请公布号 CN 113378552 A

CN 110209789 A, 2019.09.06

CN 110968775 A, 2020.04.07

CN 111107435 A, 2020.05.05

(43) 申请公布日 2021.09.10

CN 111444367 A, 2020.07.24

CN 112800254 A, 2021.05.14

(73) 专利权人 焦点科技股份有限公司

地址 210032 江苏省南京市高新开发区星

火路软件大厦A座12F

US 2021174162 A1, 2021.06.10

WO 2020190103 A1, 2020.09.24

(72) 发明人 蔡世清 郭选陵

马龙龙; 韩先培; 孙乐. 图像的文本描述方法研究综述. 中文信息学报. 2018, (第04期), 全文.

(74) 专利代理机构 南京瑞弘专利商标事务所

(普通合伙) 32249

审查员 王亚然

专利代理师 陈建和

(51) Int. Cl.

G06F 40/258 (2020.01)

G06F 40/237 (2020.01)

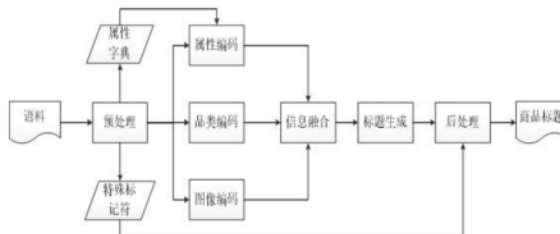
权利要求书1页 说明书4页 附图1页

(54) 发明名称

一种基于多模态GPT2模型的商品标题生成方法

(57) 摘要

基于多模态GPT2模型的商品标题生成方法, 将不同模态的商品信息融入GPT2的上下文状态中, 在理解商品内容的基础上生成商品标题: 1) 商品语音资料做预处理, 2) 采用商品信息编码模块, 分别使用ResNet和Embedding对商品图片和属性做编码, 用GPT2对商品品类名称编码, 得到GPT2网络与ResNet图像编码器、Embedding属性编码器三种不同模态信息的编码表示; 3) 采用标题生成模块; 4) 对生成的商品标题进行后处理, 识别出商品标题中的特殊标记符并替换为对应的商品属性; 完善生成后的标题文本的规格、参数信息。



1. 基于多模态GPT2模型的商品标题生成方法,其特征在于,将不同模态的商品信息融入GPT2的上下文状态中,在理解商品内容的基础上生成商品标题,具体步骤如下:步骤1:商品语料做预处理,统计属性字典和特殊标记符,构建属性字典,并且对规格类敏感信息做脱敏处理;并得到脱敏后的商品标题;步骤2:采用商品信息编码模块,分别使用ResNet和Embedding对商品图片和属性做编码,用GPT2对商品品类名称编码,得到GPT2网络与ResNet图像编码器、Embedding属性编码器三种不同模态信息的编码表示;步骤3:采用标题生成模块,将以上三种模态的商品信息编码添加到GPT2的上下文状态中,对GPT2网络与ResNet图像编码器、Embedding属性编码器三种模态的编码信息用GPT2的注意力层进行融合;用GPT2的注意力机制融合所有的上下文信息,并自回归式地生成商品标题;步骤4:对生成的商品标题进行后处理,识别出商品标题中的特殊标记符并替换为对应的商品属性;完善生成后的标题文本的规格、参数信息;所述步骤1中,处理流程具体包括如下步骤:步骤1-1:构建商品属性字典,统计所有属性中的高频属性,将商品属性的键与值拼接之后加入商品属性字典中;步骤1-2:脱敏处理,将商品属性字典中标题文本中的规格参数信息使用标记属性的特殊符号代替。

2. 根据权利要求1所述的基于多模态GPT2模型的商品标题生成方法,其特征是,所述步骤2中,需要用不同结构的子网络分别对商品不同模态的信息做编码表示,处理流程具体包括如下步骤:步骤2-1:将商品图片用ResNet子网络编码得到特征图 $I \in R^{w \times h \times s}$,其中 w 和 h 分别表示编码后的宽和高, s 表示特征图向量维度,经过一个线性变换将特征图的维度 s 变换为 d , d 表示GPT2的隐藏层维度大小,再通过形状变换将 w 和 h 两个维度合并,最后得到图片状态向量 $H_i \in R^{w \cdot h \times d}$;步骤2-2:将商品属性通过步骤一的属性字典做属性编码,再使用维度大小为 d 的Embedding子网络做向量嵌入,得到 $H_a \in R^{k \times d}$,其中 k 表示属性数量;步骤2-3:将商品标题通过GPT2得到文本表征 $H_t \in R^{1 \times d}$,其中 1 表示标题长度。

3. 根据权利要求1所述的基于多模态GPT2模型的商品标题生成方法,其特征是,所述步骤3中,将步骤2中得到的所有特征向量拼接到不同位置上 $H = [H_i; H_a; H_t]$, $[\ ; \]$ 表示拼接操作,假设当前已经生成的文本状态为 $X \in R^{n \times d}$,其中 n 表示已经生成的词语数量,则 $K = V = [H; X]$, $Q = X$,从 A 的概率分布中采样生成下一个词语;所述步骤4中,将步骤3生成的文本做进一步后处理还原出常规文本,具体操作为将生成文本中的特殊标记符,通过商品属性查表将特殊标记符替换为对应的属性值。

一种基于多模态GPT2模型的商品标题生成方法

技术领域

[0001] 本发明涉及多模态理解、文本生成技术,涉及一种通过多种模态信息(包括自然语言处理模型等)理解商品内容并自动生成商品标题的技术。

背景技术

[0002] 电子商务平台需要频繁发布商品信息,包括商品的品类、属性、标题、描述以及图片等信息,尤其以标题、描述等文本信息的编辑最为耗费人力。商品标题的文本组织需要用高度精炼的语言突出商品的特点,要求在商品内容理解的基础上准确地描述商品信息,同时平台上海量的商品信息需要保证多样性,这对商品标题发布提出了非常高的智能化要求。

[0003] 目前以GPT2为代表的文本生成模型在大量参数与海量语料的预训练基础上,可以生成非常连贯的文本,达到以假乱真的效果,但是这类模型很难控制生成的内容,而商品标题的发布需要与商品本身内容高度相关,需要很强的条件控制能力。实际上商品本身的信息是丰富的,包括品类、属性、图片多种不同模态的信息,而深度神经网络可以灵活地处理不同结构的数据,甚至融合不同模态的信息。

[0004] 本发明使用深度神经网络对商品不同模态的信息分别做编码表示,然后使用GPT2注意力机制对这些信息做交互式融合,在商品内容理解的基础上生成标题,达到严格的内容条件控制。

发明内容

[0005] 本发明目的是,克服现有技术的不足,提供一种高可控、多样性的商品标题生成算法模型。基于多模态GPT2模型对自然语音和图片等处理生成商品标题。

[0006] 为解决上述技术问题,本发明提供的技术方案是,基于多模态GPT2模型的商品标题生成方法,其特征在于,能够将不同模态的商品信息融入GPT2的上下文状态中,在理解商品内容的基础上生成相关的、高质量(符合性好,更具有接近标准的标题)的商品标题,具体包括如下步骤:

[0007] 步骤1:商品语料做预处理,构建属性字典,并且对规格类敏感信息做脱敏处理;

[0008] 步骤2:采用商品信息编码模块,分别使用ResNet和Embedding对商品图片和属性做编码,用GPT2对商品品类名称编码,得到GPT2网络与ResNet图像编码器、Embedding属性编码器三种不同模态信息的编码表示;

[0009] 步骤3:标题生成模块,将以上三种模态的商品信息编码添加到GPT2的上下文状态中,用GPT2的注意力机制融合所有的上下文信息,并自回归式地生成标题;

[0010] 步骤4:后处理,完善生成后的标题文本的规格、参数等敏感信息。

[0011] 所述步骤1中,所述处理流程具体包括如下步骤:

[0012] 步骤1-1:构建属性字典,统计所有属性中的高频属性,将属性的键与值拼接之后加入属性字典中,例如{"Color:Red", "Bound:2.4GHZ", "Size:26Inch"};

[0013] 步骤1-2:脱敏处理,将标题文本中的规格参数信息使用标记属性的特殊符号代替。

[0014] 所述步骤2中,需要用不同结构的子网络分别对商品不同模态的信息做编码表示,处理流程具体包括如下步骤:

[0015] 步骤2-1:将商品图片用ResNet子网络编码得到特征图 $I \in \mathbb{R}^{w \times h \times s}$,其中 w 和 h 分别表示编码后的宽和高, s 表示特征图向量维度,经过一个线性变换将特征图的维度 s 变换为 d , d 表示GPT2的隐藏层维度大小,再通过形状变换将 w 和 h 两个维度合并,最后得到图片状态向量 $H_i^{w \cdot h \times d}$;

[0016] 步骤2-2:将商品属性通过步骤一的属性字典做属性编码,再使用维度大小为 d 的Embedding子网络做向量嵌入,得到 $H_a^{k \times d}$,其中 k 表示属性数量;

[0017] 步骤2-3:将商品标题通过GPT2得到文本表征 $H_t^{1 \times d}$,其中 1 表示标题长度。

[0018] 所述步骤3中,将步骤2中得到的所有文本表征特征向量拼接到不同位置上 $H = [H_i; H_a; H_t]$, $[,]$ 表示拼接操作,假设当前已经生成的文本状态为 $X \in \mathbb{R}^{n \times d}$,其中 n 表示已经生成的词语数量,则 $K = V = [H; X]$, $Q = X$, $A = \text{Softmax}\left(\frac{QK^T}{d}\right)V$,从 A 的概率分布中采样生成下一个词语;

[0019] 所述步骤4中,将步骤3生成的文本做进一步后处理还原出常规文本,具体操作为将生成文本中的特殊标记符,通过商品属性查表将特殊标记符替换为对应的属性值,例如将生成的“[Tire size]All Steel Radial Truck and Bus Tire([Model NO.]+[Spec]) with Superb Wear”中的特殊标记符替换为对应的商品属性,处理后变为“20Inch All Steel Radial Truck and Bus Tire(AR5157A+12.00R20)with Superb Wear”。

[0020] 有益效果:通过商品多种不同模态的信息综合理解商品内容,从而能生成准确描述商品信息的标题文本,而利用GPT2语言模型生成更加自然、多样化的标题,达到快速发布高质量商品标题的目标。

附图说明

[0021] 图1为本发明的示例性实施例中总体流程示意图;

[0022] 图2为本发明的示例性实施例中模型结构示意图,即OpenAI预训练模型示意图;

具体实施方式

[0023] 下面结合附图和示例性实施例对本发明作进一步的说明:

[0024] 如图1所示,本发明公开一种基于多模态GPT2的商品标题生成方案,包括统一的前处理和后处理流程。

[0025] GPT是一个NLP(自然语言处理)模型,GPT-2是GPT的升级版,其最大的区别在于规模更多,训练数据更多,GPT是12层的transformer,BERT最深是24层的transformer,GPT-2则可以是48层,其训练数据是一个WebText的数据集,该数据集做了一些简单的数据清理,并且覆盖领域十分广阔。

[0026] 步骤1:对语料做预处理,统计属性字典和特殊标记符,并得到脱敏后的商品标题,;

[0027] 步骤2:对商品内容进行编码,包括用ResNet对图片编码,用GPT2对品类名称编码,以及用Embedding对属性编码;

[0028] 步骤3:对三种模态的编码信息用GPT2的注意力层进行融合;对GPT2网络与ResNet图像编码器、Embedding属性编码器三种模态的编码信息用GPT2的注意力层进行融合;

[0029] 步骤4:用GPT2对融合后的状态信息生成商品标题;

[0030] 步骤5:对生成的标题进行后处理,识别出标题中的特殊标记符并替换为对应的商品属性;

[0031] 多模态GPT2的商品标题生成模型,可以综合商品的各个模态信息,全面理解商品内容,生成能够描述商品特点的高质量标题。

[0032] 多模态GPT2网络模型,具体包含GPT2网络与ResNet图像编码器、Embedding属性编码器。

[0033] 如图2所示,本发明公开一种多模态GPT2网络模型,具体包含GPT2网络与ResNet图像编码器、Embedding属性编码器。

[0034] 商品的品类名称(Catalog)直接作为文本输入GPT2。所述GPT2网络使用图2的OpenAI预训练模型,包含Word Embedding(词语嵌入层)、Position Embedding(位置编码层)、N个Transformer块、一个Linear(线性层)以及Softmax输出激活函数,其中每个Transformer块包括一个Masked Multi-Head Attention(多头注意力层)与Feed Forward(前馈层),并且在多头注意力层与前馈层后面分别连接一个残差层与LayerNorm层。

[0035] 所述ResNet图像编码器采用ResNet50预训练骨干网络再接一个线性变换层,图像统一size为 224×224 ,ResNet50输出为 $7 \times 7 \times 2048$,再用线性变换层得到输出 $7 \times 7 \times 1024$ 。

[0036] 所述Embedding属性编码器包括D个属性向量,D表示属性字典的大小,向量的维度也是1024维。因推荐系统分召回和排序,召回为将找出可能语词的候选集,排序是对候选集按照某种指标程度进行排序,最终得出给用户推荐的结果。在推荐系统的召回阶段,需要对每个用户和每个被推荐语词做数学层面的表示,通过向量即Embedding表示。具体为Embedding属性编码器。

[0037] 文本Embedding方法,也是目前最流行的方案是动态词表征算法,比如Bert、ELMo、GPT,这类算法会考虑文本上下文。

[0038] 在网站上获得图像和网文描述商品信息的标题文本,输入的量有200M,先将商品的品类名称(Catalog)直接作为文本输入GPT2后,品类名称处理成商品属性字典,统计所有属性中的高频属性,将商品属性的键与值拼接之后加入商品属性字典中,将商品属性字典中标题文本中的规格参数信息使用标记属性的特殊符号代替,例如原始文本为“20Inch All Steel Radial Truck and Bus Tire (AR5157A+12.00R20)with Superb Wear”,处理后为“[Tire size]All Steel Radial Truck and Bus Tire ([Model NO.] + [Spec])with Superb Wear”,并将统计得到的特殊符号如本例中的“[Tire size]”等添加到GPT2词表中。

[0039] 将商品图片用ResNet子网络编码得到特征图 $I \in \mathbb{R}^{w \times h \times s}$,其中w和h分别表示编码后的宽和高,s表示特征图向量维度,经过一个线性变换将特征图的维度s变换为d,d表示GPT2的隐藏层维度大小,再通过形状变换将w和h两个维度合并,最后得到图片状态向量 $H_1^{w \cdot h \times d}$;

[0040] 再使用维度大小为d的Embedding子网络做向量嵌入,得到 $H_a^{k \times d}$,其中k表示属性数量;

[0041] 将商品标题通过GPT2得到文本表征 $H_t^{1 \times d}$,其中1表示标题长度。

[0042] 将步骤2中得到的所有特征向量拼接到不同位置上 $H=[H_i;H_a;H_t]$,[,]表示拼接操作,假设当前已经生成的文本状态为 $X \in R^{n \times d}$,其中n表示已经生成的词语数量,则 $K=V=[H;X]$, $Q=X$, $A=Softmax\left(\frac{QK^T}{d}\right)V$,从A的概率分布中采样生成下一个词语;

[0043] 将步骤3生成的文本做进一步后处理还原出常规文本,具体操作为将生成文本中的特殊标记符,通过商品属性查表将特殊标记符替换为对应的属性值,例如将生成的“[Tire size]All Steel Radial Truck and Bus Tire([Model NO.]+[Spec])with Superb Wear”中的特殊标记符替换为对应的商品属性,处理后变为“20Inch All Steel Radial Truck and Bus Tire(AR5157A+12.00R20)with Superb Wear”。

[0044] 所述GPT2网络使用图2的OpenAI预训练模型,包含Word Embedding(词语嵌入层)、Position Embedding(位置编码层)、N个Transformer块、一个Linear(线性层)以及Softmax输出激活函数,其中每个Transformer块包括一个Masked Multi-Head Attention(多头注意力层)与Feed Forward(前馈层),并且在多头注意力层与前馈层后面分别连接一个残差层与LayerNorm层。商品的品类名称(Catalog)直接作为文本输入GPT2。

[0045] 所述ResNet图像编码器采用ResNet50预训练骨干网络再接一个线性变换层,图像统一size为 224×224 ,ResNet50输出为 $7 \times 7 \times 2048$,再用线性变换层得到输出 $7 \times 7 \times 1024$ 。

[0046] 所述Embeddi。

[0047] 将图像编码输出、属性编码输出与商品品类的表示在多头注意力层做融合,获得商品的综合表征状态,再用GPT2完成标题的生成操作。即应用三种编码得到结果并融合。而利用GPT2语言模型生成更加自然、多样化的标题,达到快速发布高质量商品标题的目标。

[0048] 本发明主要用于提供一种在电子商务场景下自动化、智能化发布商品标题的模型,在添加少量额外编码器的基础上充分利用GPT2预训练语言模型,可以生成高质量、多样化的商品标题描述。

[0049] 以上实施例不以任何方式限定本发明,凡是对以上实施例以等效变换方式做出的其它改进与应用,都属于本发明的保护范围。

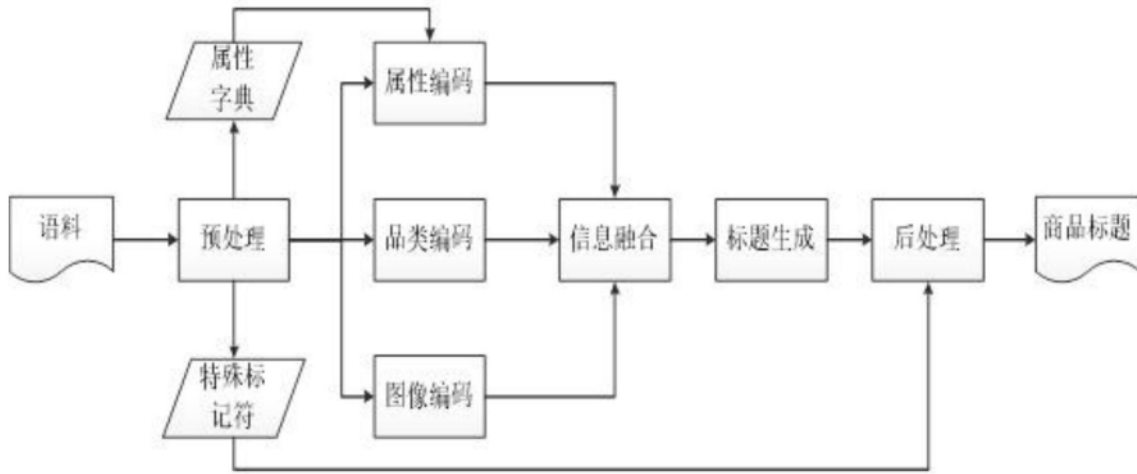


图1

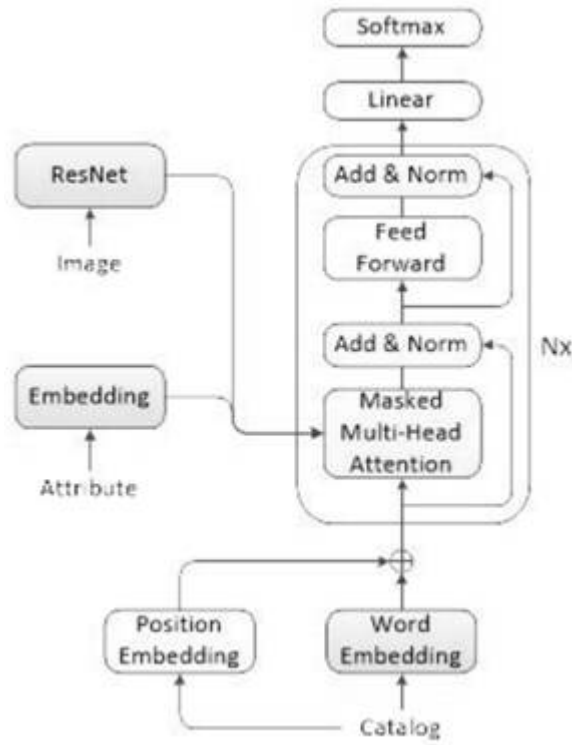


图2