



(12) 发明专利申请

(10) 申请公布号 CN 113868459 A

(43) 申请公布日 2021. 12. 31

(21) 申请号 202110712040.9

G06N 3/04 (2006.01)

(22) 申请日 2021.06.25

G06N 3/08 (2006.01)

(71) 申请人 之江实验室

地址 310023 浙江省杭州市余杭区文一西路1818号人工智能小镇10号楼

申请人 复旦大学

(72) 发明人 魏忠钰 李泽君

(74) 专利代理机构 杭州浙科专利事务所(普通合伙) 33213

代理人 孙孟辉

(51) Int. Cl.

G06F 16/58 (2019.01)

G06F 16/33 (2019.01)

G06F 40/211 (2020.01)

G06F 40/289 (2020.01)

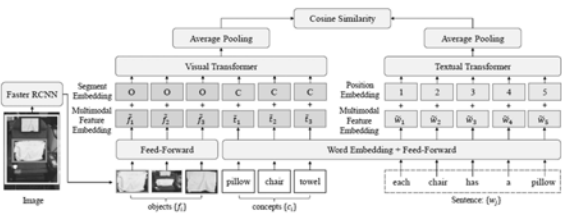
权利要求书2页 说明书13页 附图3页

(54) 发明名称

模型训练方法、跨模态表征方法、无监督图像文本匹配方法及装置

(57) 摘要

本发明的目的是提供一种模型训练方法、跨模态表征方法、无监督图像文本匹配方法及装置,所述方法包括:计算训练文档中图片与句子的两两相似度值;基于所述相似度值,确定正样本对集和负样本对集;其中,所述正样本对集中有预设数量的正样本对;所述负样本对集中有预设数量的负样本对;所述正样本对集和所述负样本对集用于进一步训练所述模型,直至预设数量的所述正样本对的平均相似度值大于预设数量的所述负样本对的平均相似度值,且两者差值符合预设条件。上述实施方式可以减小采样的偏差,以更好的训练模型来对图片和句子进行匹配。



1. 一种模型训练方法,其特征在于,应用于无监督图像文本匹配模型,所述方法包括:
计算训练文档中图片与句子的两两相似度值;
基于所述相似度值,确定正样本对集和负样本对集;其中,所述正样本对集中有预设数量的正样本对;所述负样本对集中有预设数量的负样本对;所述正样本对集和所述负样本对集用于进一步训练所述模型,直至预设数量的所述正样本对的平均相似度值大于预设数量的所述负样本对的平均相似度值,且两者差值符合预设条件。
2. 如权利要求1所述方法,其特征在于,在确定正样本对集和负样本对集的步骤中包括:
在所述图片与所述句子来自同一文档的情况下:
对应于每个所述图片,将该图片与其相似度值最高的所述句子放入候选正样本对集;
将所述候选正样本对集中相似度值最高的预设数量的所述正样本对放入所述正样本对集;
和/或,
对应于每个所述句子,将该句子与其相似度值最高的所述图片放入候选正样本对集;
将所述候选正样本对集中相似度值最高的预设数量的所述正样本对放入所述正样本对集。
3. 如权利要求1所述方法,其特征在于,在确定正样本对集和负样本对集的步骤中包括:
在所述图片与所述句子来自同一文档的情况下:
对应于每个所述图片,将该图片与其相似度值最低的所述句子放入候选负样本对集;
将所述候选负样本对集中相似度值最低的预设数量的所述负样本对放入所述负样本对集;
和/或,
对应于每个所述句子,将该句子与其相似度值最低的所述图片放入候选负样本对集;
将所述候选负样本对集中相似度值最低的预设数量的所述负样本对放入所述负样本对集。
4. 如权利要求1所述方法,其特征在于,在确定正样本对集和负样本对集的步骤中包括:
在确定所述负样本对集的情况时,所述图片与所述句子来自不同文档;
对应于每个所述图片,将该图片与其相似度值最高的所述句子放入候选负样本对集;
将所述候选负样本对集中相似度值最高的预设数量的所述负样本对放入所述负样本对集;
和/或,
对应于每个所述句子,将该句子与其相似度值最高的所述图片放入候选负样本对集;
将所述候选负样本对集中相似度值最高的预设数量的所述负样本对放入所述负样本对集。
5. 如权利要求1所述方法,其特征在于,所述训练文档为残次文档即该文档中部分句子和/或图片被遮盖。
6. 一种模型训练装置,其特征在于,应用于无监督图像文本匹配模型,所述装置包括:

计算单元,所述计算单元用于计算训练文档中图片与句子的两两相似度值;

采样单元,所述采样单元用于基于所述相似度值,确定正样本对集和负样本对集;其中,所述正样本对集中有预设数量的正样本对;所述负样本对集中有预设数量的负样本对;所述正样本对集和所述负样本对集用于进一步训练所述模型,直至预设数量的所述正样本对的平均相似度值大于预设数量的所述负样本对的平均相似度值,且两者差值符合预设条件。

7. 一种跨模态表征方法,其特征在于,应用于无监督图像文本匹配模型的训练或无监督图像文本匹配,所述方法包括:

根据获取的目标文档的图片信息,获取对应图片的图片表征信息,所述图片表征信息包括该图片信息对应的至少一个对象,以及概念信息;所述概念信息为根据所述对象预测出的标签;

根据获取的所述目标文档的句子信息,获取对应句子的句子表征信息;

将所述图片表征信息和所述句子表征信息进行匹配,得到对应于该图片与该句子的匹配值;所述匹配值用于表征该图片与该句子的匹配度。

8. 如权利要求7所述方法,其特征在于,所述概念信息和所述句子信息的表征信息共享嵌入层。

9. 一种无监督图像文本匹配方法,其特征在于,所述方法包括:

基于如权利要求1-5的任一模型训练方法,训练无监督图像文本匹配模型;

基于所述无监督图像文本匹配模型,获取目标文档的图片对应的图片表征信息,所述图片表征信息包括该图片信息对应的至少一个对象,以及概念信息;所述概念信息为根据所述对象预测出的标签;

基于所述无监督图像文本匹配模型,获取所述目标文档对应句子的句子表征信息;

将所述图片表征信息和所述句子表征信息进行匹配,得到对应于该图片与该句子的匹配值;所述匹配值用于表征该图片与该句子的匹配度。

10. 一种无监督图像文本匹配装置,其特征在于,所述装置包括:

训练单元,所述训练单元用于训练无监督图像文本匹配模型;

无监督图像文本匹配单元,所述无监督图像文本匹配单元用于基于所述训练单元,匹配目标文档的中的图片与句子;所述无监督图像文本匹配单元至少包括:

第一提取单元,所述第一提取单元用于获取所述目标文档的图片对应的图片表征信息,所述图片表征信息包括该图片信息对应的至少一个对象,以及概念信息;所述概念信息为根据所述对象预测出的标签;

第二提取单元,所述第二提取单元用于获取所述目标文档对应句子的句子表征信息;

匹配单元,所述匹配单元用于将所述图片表征信息和所述句子表征信息进行匹配,得到对应于该图片与该句子的匹配值;所述匹配值用于表征该图片与该句子的匹配度。

模型训练方法、跨模态表征方法、无监督图像文本匹配方法及装置

技术领域

[0001] 本发明涉及计算机领域,尤其涉及一种模型训练方法、跨模态表征方法、无监督图像文本匹配方法及装置。

背景技术

[0002] 图像-文本匹配是视觉和语言领域的基本问题之一,主要目标是学习对齐两种模式的语义空间。以往关于图像-文本匹配的研究主要是监督的,需要大量的标注图像-句子对。但是在有监督的图像-文本匹配模型中,图像和句子的标记对获取代价昂贵。

[0003] 目前研究探讨利用文档级结构信息抽取正、负实例进行模型训练,但存在抽样偏差,导致模型无法区分语义相似度高的实例。

发明内容

[0004] 本说明书实施方式的目的是提供一种模型训练方法、跨模态表征方法、无监督图像文本匹配方法及装置。本说明书实施方式提供的新的训练方法(采样方法),可以减小采样的偏差。进一步,本说明书实施方式提供的跨模态表征方法、无监督图像文本匹配方法基于Transformer的模型来识别更为复杂的语义关联,该模型可以为每个多模态文档隐式地构建了一个图的结构,构建同一篇文档内句子和图片的表征学习间的桥梁,可以更好的匹配目标文档的图片句子。

[0005] 为实现上述目的,本说明书实施方式提供了一种模型训练方法,应用于无监督图像文本匹配模型,所述方法包括:计算训练文档中图片与句子的两两相似度值;基于所述相似度值,确定正样本对集和负样本对集;其中,所述正样本对集中有预设数量的正样本对;所述负样本对集中有预设数量的负样本对;所述正样本对集和所述负样本对集用于进一步训练所述模型,直至预设数量的所述正样本对的平均相似度值大于预设数量的所述负样本对的平均相似度值,且两者差值符合预设条件。

[0006] 在一个实施方式中,在确定正样本对集和负样本对集的步骤中包括:在所述图片与所述句子来自同一文档的情况下:对应于每个所述图片,将该图片与其相似度值最高的所述句子放入候选正样本对集;将所述候选正样本对集中相似度值最高的预设数量的所述正样本对放入所述正样本对集;和/或,对应于每个所述句子,将该句子与其相似度值最高的所述图片放入候选正样本对集;将所述候选正样本对集中相似度值最高的预设数量的所述正样本对放入所述正样本对集。

[0007] 在一个实施方式中,在确定正样本对集和负样本对集的步骤中包括:在所述图片与所述句子来自同一文档的情况下:对应于每个所述图片,将该图片与其相似度值最低的所述句子放入候选负样本对集;将所述候选负样本对集中相似度值最低的预设数量的所述负样本对放入所述负样本对集;和/或,对应于每个所述句子,将该句子与其相似度值最低的所述图片放入候选负样本对集;候选负样本对集将所述候选负样本对集中相似度值

最低的预设数量的所述负样本对放入所述负样本对集。

[0008] 在一个实施方式中,在确定正样本对集和负样本对集的步骤中包括:在确定所述负样本对集的情况时,所述图片与所述句子来自不同文档;对应于每个所述图片,将该图片与其相似度值最高的所述句子放入候选负样本对集;将所述候选负样本对集中相似度值最高的预设数量的所述负样本对放入所述负样本对集;和/或,对应于每个所述句子,将该句子与其相似度值最高的所述图片放入候选负样本对集;将所述候选负样本对集中相似度值最高的预设数量的所述负样本对放入所述负样本对集。

[0009] 在一个实施方式中,所述训练文档为残次文档即该文档中部分句子和/或图片被遮盖。

[0010] 本说明书实施方式还提供了一种模型训练装置,应用于无监督图像文本匹配模型,所述装置包括:计算单元,所述计算单元用于计算训练文档中图片与句子的两两相似度值;采样单元,所述采样单元用于基于所述相似度值,确定正样本对集和负样本对集;其中,所述正样本对集中有预设数量的正样本对;所述负样本对集中有预设数量的负样本对;所述正样本对集和所述负样本对集用于进一步训练所述模型,直至预设数量的所述正样本对的平均相似度值大于预设数量的所述负样本对的平均相似度值,且两者差值符合预设条件。

[0011] 本说明书实施方式还提供了一种跨模态表征方法,应用于无监督图像文本匹配模型的训练或无监督图像文本匹配,所述方法包括:根据获取的目标文档的图片信息,获取对应图片的图片表征信息,所述图片表征信息包括该图片信息对应的至少一个对象,以及概念信息;所述概念信息为根据所述对象预测出的标签;根据获取的所述目标文档的句子信息,获取对应句子的句子表征信息;将所述图片表征信息和所述句子表征信息进行匹配,得到对应于该图片与该句子的匹配值;所述匹配值用于表征该图片与该句子的匹配度。

[0012] 在一个实施方式中,所述概念信息和所述句子信息的表征信息共享嵌入层。

[0013] 在一个实施方式中,一种无监督图像文本匹配方法,所述方法包括:基于如上任一模型训练方法,训练无监督图像文本匹配模型;基于所述无监督图像文本匹配模型,获取目标文档的图片信息的图片表征信息,所述图片表征信息包括该图片信息对应的至少一个对象,以及概念信息;所述概念信息为根据所述对象预测出的标签;基于所述无监督图像文本匹配模型,获取所述目标文档对应句子的句子表征信息;将所述图片表征信息和所述句子表征信息进行匹配,得到对应于该图片与该句子的匹配值;所述匹配值用于表征该图片与该句子的匹配度。

[0014] 本说明书实施方式还提供了一种无监督图像文本匹配装置,所述装置包括:训练单元,所述训练单元用于训练无监督图像文本匹配模型;无监督图像文本匹配单元,所述无监督图像文本匹配单元用于基于所述训练单元,匹配目标文档中的图片与句子;所述无监督图像文本匹配单元至少包括:第一提取单元,所述第一提取单元用于获取所述目标文档的图片对应的图片表征信息,所述图片表征信息包括该图片信息对应的至少一个对象,以及概念信息;所述概念信息为根据所述对象预测出的标签;第二提取单元,所述第二提取单元用于获取所述目标文档对应句子的句子表征信息;匹配单元,所述匹配单元用于将所述图片表征信息和所述句子表征信息进行匹配,得到对应于该图片与该句子的匹配

值;所述匹配 值用于表征该图片与该句子的匹配度。

[0015] 由以上本说明书实施方式提供的技术方案可见,本说明书实施方式的目的是 提供一种模型训练方法、跨模态表征方法、无监督图像文本匹配方法及装置。本 说明书实施方式的提供的新的训练方法(采样方法),可以减小采样的偏差;进 一步,本说明书实施方式的跨模态表征方法、无监督图像文本匹配方法基于Transformer的模型来识别更为复杂的语义关联,该模型为每个多模态文档隐式 地构建了一个图的结构,构建了同一篇文档内句子和图片的表征学习间的桥梁。实验的结果也证明了本说明书提供的方法和装置可以有有效的减小偏差并且进一 步获得了更好的跨模态表征。

附图说明

[0016] 为了更清楚地说明本说明书实施方式或现有技术中的技术方案,下面将对实 施方式或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描 述中的附图仅仅是本说明书中记载的一些实施方式,对于本领域普通技术人员来 讲,在不付出创造性劳动性的前提下,还可以根据这些附图获得其他的附图。

[0017] 图1是本说明书提供的一种图像-句子匹配的不同设置说明示意图;

[0018] 图2是本说明书提供的一个示例的示意图;

[0019] 图3是本说明书提供的一个偏差示意图;

[0020] 图4是本说明书提供的一个三个部分的采样和训练目标示意图;

[0021] 图5是本说明书提供的一个总的模型结构示意图;

[0022] 图6是本说明书提供的一个有监督-无监督比较示意图。

具体实施方式

[0023] 下面将结合本说明书实施方式中的附图,对本说明书实施方式中的技术方案 进行清楚、完整地描述,显然,所描述的实施方式仅仅是本说明书一部分实施方 式,而不是全部的实施方式。基于本说明书中的实施方式,本领域普通技术人员 在没有作出创造性劳动前提下所获得的所有其他实施方式,都应当属于本申请保 护的范围。

[0024] 本说明书提供的一种模型训练方法,应用于无监督图像文本匹配模型,所述 方法可以包括以下步骤。

[0025] 步骤S10:计算训练文档中图片与句子的两两相似度值。

[0026] 步骤S12:基于所述相似度值,确定正样本对集和负样本对集;其中,所述 正样本对集中有预设数量的正样本对;所述负样本对集中有预设数量的负样本对; 所述正样本对集和所述负样本对集用于进一步训练所述模型,直至预设数量的所 述正样本对的平均相似度值大于预设数量的所述负样本对的平均相似度值,且两 者差值符合预设条件。

[0027] 为了方便本申请说明,请参阅图1。图片-句子的匹配一直是跨模态领域的基 础任务,其根本的目的是对其视觉和文本的语义空间。图1为图像-句子匹配的 不同设置说明:(a)多模态语义空间,(b)监督对齐,(c)无监督跨文档目标,(d)本 说明书提供的模型训练方法(文档内部目标)。深色链接表示匹配的正配对,浅 色链接表示负配对;实链接表示标注标签即监督对齐,虚线表示无监督方法检测 到的伪标签,即深色/浅色虚线代表无监督方法选出的(伪)正/负样本对;虚线 圈表示文档中的图像集和句子集。左右两侧点分别代

表图片/句子,

[0028] 如(a)所示,两个模态之间本身存在着语义空间上的差异,对其的常见方法 是通过有监督的标签拉近匹配的样本对,如(b)所示。在无监督的环境下,最大 的挑战即为如何选择出想要拉近的正样本对和远离的负样本对。

[0029] 如(c)所示,(c)表征一种训练方法,即通过文档的图片句子共现信息,通过 拉近句子集和图片集的方式来进行训练,其中,同文档内语义近似的句子-图片 对被看作正样本,而跨文档间的句子-图片对被看作负样本,如(c)所示,这样的 方法没有考虑到文档内部语义相似度更高的负样本,其选出的负样本与真实的负 样本分布存在着偏差。

[0030] 如(d)所示,(d)表征说明书提供的模型训练方法,引入了更多同一文档内部 的正负样本对来帮助训练。

[0031] 请参阅图2、图3。图2为在训练和评估的正样本和负样本说明:深色/浅色链接 是文档内考虑的负样本/正样本,左右两幅图中间的链接为文档间训练考虑的负 样本。图3为在文档间训练和文档内训练过程中,对于同一句子,预先训练的 ground-truth匹配图像和负图像的CNN特征之间L2距离的分布。

[0032] 图2显示了一个示例。训练的正样本和负样本对更容易区分(书和马),而测 试环境中的负样本中的图像和句子在语义上具有高度的相关性。进一步在图3中 展示了导致这种抽样策略的偏差的证据。对于VIST-DII数据集中的每一个句子, 计算其对应图像的CNN特征与训练过程和评估过程中考虑的负面图像之间的L2 距离,然后将这两个分布可视化。这两个分布之间的差异是显著的,训练样本的 距离通常较大。这样的抽样偏差使得训练过的模型很难学习到良好的表示方法来 确定相似图像和句子之间的对应关系。

[0033] 为了缓解采样偏差的问题,在同一文档中进一步区分正样本和负样本,如图 1(d)所示。在实践中,从一个文档中语义相似度最低的对中抽取负样本。区分文 档中相似的图像和句子,需要主干模型具有使用细粒度信息学习跨模态表示的能 力。考虑图2右侧的文档,捕捉对象级的细节(如“dirt”和“grass”)是区分这两 句话和图像的必要条件。进一步探索在文档中建立概念模型,以桥接图像和句子 的语义。本说明书实施方式提出从图像中提取概念,并建立一个由图像、句子和 概念组成的文档内图结构。一个基于Transformer的模型被用来对这些隐式依赖进 行建模,并用上下文编码的细粒度特征来表示图像和句子。

[0034] 在本实施方式中,计算训练文档中图片与句子的两两相似度值。所述相似度 值用于表征对应图片与句子的相似度。

[0035] 在本实施方式中,基于所述相似度值,确定正样本对集和负样本对集;其中, 所述正样本对集中有预设数量的正样本对;所述负样本对集中有预设数量的负样 本对;所述正样本对集和所述负样本对集用于进一步训练所述模型,直至预设数 量的所述正样本对的平均相似度值大于预设数量的所述负样本对的平均相似度 值,且两者差值符合预设条件。具体的,可以包括文档内部的训练目标 (intra-document objective)、跨文档训练目标 (cross-document objective) 和次跨文档 训练目标(dropout sub-document objective)。

[0036] 在本实施方式中,文档内部的训练目标(intra-document objective)具体的,例 如,对于一个包含了N个句子和M张图片的文档,基于当前的匹配模型计算两 两之间的相似度,采样正样本对(图片-句子对)时:从文本端出发,对于每一 个句子,基于当前模型计算

出的相似度,都会选择出与其相似度最高的一张图片 作为候选,与该句子一起作为一个候选的样本对。最后在这N个候选的样本对 中选择相似度最高的K对作为正样本对。从视觉端出发,对于每一个图片,基 于当前模型计算出的相似度,都会选择出与其相似度最高的一条句子,与该图片 一起作为一个候选的样本对。最后在这M个候选的样本对中选择相似度最高的 K对作为正样本对。对于一个包含了N个句子和M张图片的文档,采样负样本 对(图片-句子对)时:从文本端出发,对于每一个句子,基于当前模型计算出 的相似度,都会选择出与其相似度最低的一张图片作为候选,与该句子一起作为 一个候选的样本对。最后在这N个候选的样本对中选择相似度最低的K对作为 负样本对。从视觉端出发,对于每一个图片,基于当前模型计算出的相似度,都 会选择出与其相似度最低的一条句子,与该图片 一起作为一个候选的样本对。最 后在这M个候选的样本对中选择相似度最低的K对作为负样本对。该部分的训 练目标(intra-document objective)即要求这个部分得到的正样本的平均相似度要 高于负样本的平均相似度一定的值 $\alpha/2$ (基于铰链损失定义)。

[0037] 在本实施方式中,跨文档训练目标(cross-document objective)具体的,例如,该目标的得到正样本的方法与上一目标中的方法一样,为来自同一文档的句子图 片对;而这一部分得到的负样本为来自不同文档的图片-句子对,具体的方法为: 基于当前的匹配模型计算来自两个不同文档的多对图片句子之间的相似度,使用 TK函数从中选择K对相似度最高的样本对作为负样本。这一部分的训练目标 (cross-document objective)同样也是要求这个部分得到的正样本的平均相似度 要高于负样本的平均相似度一定的值 α (基于铰链损失定义)。

[0038] 在本实施方式中,次跨文档训练目标(dropout sub-document objective)具体的, 该部分的训练目标与第二部分类似,不过这里进行正样本采样的时候,考虑的文 档为残次文档(部分图片和句子被遮盖),同样是在残次文档里用TK采样得到 正样本对,在来自不同文档的图片句子对中用TK得到负样本对。这一部分的训 练目标(dropout sub-document objective)同样也是要求这个部分得到的正样本的 平均相似度要高于负样本的平均相似度一定的值 $\alpha/2$ (基于铰链损失定义)。

[0039] 上述三个训练目标都要求多个正样本对之间的平均相似度要高于多个负样 本对 的平均相似度,多个样本对之间的平均相似度可以被看作是文档级别的相似 度(存在于 正/负/残次文档中),在后续的具体说明里,这三个训练目标的定义都 是基于文档级别的。

[0040] 请参阅图4,图4为三个部分的采样和训练目标示意。文档中的线条代表句 子,不同的深浅颜色对应不同的文档。“+”表示TK函数,“-”表示NegTK函

[0041] ^

[0042] 数。 \hat{S} 是由相同的骨干模型计算,和所有的 \hat{S} 文档级别的相似之处。虚线表示铰链损失函数的负输入。

[0043] 具体的,跨文档训练目标(cross-document objective)。该目标是基于这样的假设:共现的图像集/句子集对比非共现的图像集/句子集对更相似。通过结合非共现 的图像集/句子集对构造负文档,单个正文档的目标可以用带有铰链损失的硬负 挖掘来表征:

$$\mathcal{L}_c(S_i, V_i) = \max_{j \neq i} h_{\alpha}(\text{sim}(S_i, V_i), \text{sim}(S_i, V_j)) + \max_{j \neq i} h_{\alpha}(\text{sim}(S_i, V_i), \text{sim}(S_j, V_i)) \quad (1)$$

[0044]

[0045] 在同一个批量里考虑 i, j , 代表 i 和 j 需要在整个批量里进行遍历。 $h_\alpha(m, n) = \max(0, n - m + \alpha)$ 是铰链损失函数与 α 的优势, sim 是一个相似度函数来计算相似性的图像集/句子集对之间的关联矩阵 \hat{M} 预测映射到的一个实数, 该函数将选择代表图像-句子对根据一个特定的标准, 和计算的平均相似性选择对文档级相似。在本实施方式中, 使用一个函数TK:

$$[0046] \quad \mathbb{R}^{|S| \times |V|} \mapsto \mathbb{R}$$

[0047] 这里, 根据当前预测的相似度选择 k 条最有可能的句子到图像和图像到句子的边, 然后计算所选对的平均相似度作为文档级相似度。该程序对应于公式:

$$[0048] \quad \text{sim}(S_i, V_j) = \hat{S}_{i,j}^c = \text{TK}(\hat{M}_{i,j}^c)$$

[0049] 文档内部的训练目标(intra-document objective)。其假设为同一篇文档内部的语义近似的图片句子对之间的相似度也要高于内部语义相差较远的图片句子对间的相似度, 高于一定的值, 在此目标下采样出的正样本为来自同一文档的语义较为近似的句子-图片对; 负样本为来自同一文档的语义相差较远的图片句子对。

[0050] 类似于“TK”度量一个文档内的正相似度(即预测出最可能匹配的图片句子对之间的相似度), NegTK 函数度量了一个文档内的负相似度(即预测出最不可能匹配的图片句子间的相似度)。正 NegTK 将首先选择 k 最不可能 sentence-to-image 和 image-to-sentence 边缘基于当前的预测相似, 然后计算平均相似性选择对文档级别的“负面”相似。通过文档级“正”相似度和“负”相似度之间的铰链缺失来描述这个文档内部目标:

$$\hat{S}_i^{pos} = \text{TK}(\hat{M}_i)$$

$$[0051] \quad \hat{S}_i^{neg} = \text{NegTK}(\hat{M}_i)$$

$$L_{intra}(S_i, V_i) = h_{\frac{\alpha}{2}}(\hat{S}_i^{pos}, \hat{S}_i^{neg}) \quad (2)$$

[0052]

[0053]

[0054] 其中, $h(\cdot)$ 的定义与方程1相同, 但 $\text{margin}_{\frac{\alpha}{2}}$ 较小。 NegTK 可以通过等价的定义有效地实现: $\text{NegTK}(\hat{M}_i) = -\text{TK}(-\hat{M}_i)$ 。

[0055] 实质上, 添加这个互补目标相当于将预测相似度较低的文档内图像-句子对作为负样本。一般情况下, 几乎不可能所有的图像-句子对都具有语义关联, 该策略将有很高的概率选择那些在ground truth中没有边缘的图像-句子。这里的边缘即代表匹配关系, 没有边缘代表不匹配的图片-句子对。

[0056] 次跨文档训练目标(dropout sub-document objective)。其假设为即使一篇文档将其随机的遮盖住部分的句子/图片, 剩下的残缺文档内的句子集和图片集间的相似度也要高于跨文档间的图片集-句子集间的相似度。在此目标下采样出的正样本为来自同一“残次”文档的语义较为近似的句子-图片对; 负样本为来自不同文档的语义近似的图片句子对。

[0057] 在使用TK时, 只有 $2k$ 个最可能的图像句子对被认为是正样本。除了选定的 $2k$ 条边

之外,根据文档的组成,其他图像-句子对之间可能存在(较弱的)语义关联。

[0058] 利用这些信息,引入互补跨文档的另一个目的,假设即使一些句子和图片在文档中,残次文档内的句子图片间的文档级别相似度仍然会高于负文档,只是差距相对来说更小一些。

[0059] 对于单个正向文档,通过随机删除一定百分比($1-p_{sub}$)的句子和图片来构建 dropout子文档:

$$[0060] \quad S_i^{sub} = \text{Uniform}(S_i, n = \lfloor p_{sub} \times |S_i| \rfloor)$$

$$[0061] \quad V_i^{sub} = \text{Uniform}(V_i, n = \lfloor p_{sub} \times |V_i| \rfloor) \quad (3)$$

[0062] 式中 $\text{Uniform}(A, n=n_A)$ 表示从A中均匀抽取 n_A 样本而无放回抽样的函数。然后用类似 \mathcal{L}_c 的形式来表征目标:

$$[0063] \quad \begin{aligned} \mathcal{L}_{cross}^{sub}(S_i, V_i) = & \max_{i \neq j} h_{\frac{\alpha}{2}}(\text{sim}(S_i^{sub}, V_i^{sub}), \text{sim}(S_i, V_j)) \\ & + \max_{i \neq j} h_{\frac{\alpha}{2}}(\text{sim}(S_i^{sub}, V_i^{sub}), \text{sim}(S_j, V_i)) \end{aligned} \quad (4)$$

[0064] 在文档一级,增加这一补充的目标意味着建立更正文档;在图像-句子对层次上,它相当于一种新的采样策略:从文档中选择额外的正图像-句子对,而不考虑某些图像和句子。但是这些正样本被认为是较弱的,应该有一个较小的差距。

[0065] 结合3个目标,以总损失,为一个单一的正文档,为:

$$[0066] \quad \mathcal{L}(S_i, V_i) = \mathcal{L}_c(S_i, V_i) + \mathcal{L}_{intra}(S_i, V_i) + \mathcal{L}_{cross}^{sub}(S_i, V_i) \quad (5)$$

[0067] 在本实施方式中,根据所述正、负配对样本,训练所述无监督图像文本匹配模型;其中,所述无监督图像文本匹配模型用于匹配目标文档中的图片与句子。具体的,例如,为了更好地识别同一文档内更加复杂的句子图片语义匹配模式,本说明书提供的匹配模式可以使用更加细粒度的表征学习方法,基于Transformer的模型,在其中为每个文档的句子图片间隐性建模了一个图,来帮助获得更好的跨模态表征。

[0068] 通过本实施提供的训练方法,可以缓解样本偏差问题。从文档内对中选择负样本和附加正样本的策略,形成了一个新的跨模态表示学习目标。为了识别文档内部样本中的复杂模式,本实施提供一个基于Transformer的模型来捕获细粒度特征,并将概念集成到我们的模型中,以桥接文档中多模态数据的表示学习。在多图像、多句子文档的多模型链接预测任务上对本说明书提供的方法进行了评估。实验结果表明,本说明书提供的方法能够有效地消除偏见,并能更好地学习文档中的多模态表示。

[0069] 在一个实施方式中,在确定正样本对集和负样本对集的步骤中包括:在所述图片与所述句子来自同一文档的情况下:对应于每个所述图片,将该图片与其相似度值最高的所述句子放入候选正样本对集;将所述候选正样本对集中相似度值最高的预设数量的所述正样本对放入所述正样本对集;和/或,对应于每个所述句子,将该句子与其相似度值最高的所述图片放入候选正样本对集;将所述候选正样本对集中相似度值最高的预设数量的所述正样本对放入所述正样本对集。

[0070] 该实施方式对应于文档内部的训练目标 (intra-document objective) 中确定正样本对集的步骤, 此处不做具体赘述。

[0071] 在一个实施方式中, 在确定正样本对集和负样本对集的步骤中包括: 在所述 图片与所述句子来自同一文档的情况下: 对应于每个所述图片, 将该图片与其相似 度值最低的所述句子放入候选负样本对集; 将所述候选负样本对集中相似度值 最低的预设数量的所述负样本对放入所述负样本对集; 和/或, 对应于每个所述 句子, 将该句子与其相似度值最低的所述图片放入候选负样本对集; 候选负样本 对集将所述候选负样本对集中相似度值最低的预设数量的所述负样本对放入所 述负样本对集。

[0072] 该实施方式对应于文档内部的训练目标 (intra-document objective) 中确定负样本对集的步骤, 此处不做具体赘述。

[0073] 在一个实施方式中, 在确定正样本对集和负样本对集的步骤中包括: 在确定 所述负样本对集的情况时, 所述图片与所述句子来自不同文档; 对应于每个所述 图片, 将该图片与其相似度值最高的所述句子放入候选负样本对集; 将所述候选 负样本对集中相似度值最高的预设数量的所述负样本对放入所述负样本对集; 和 /或, 对应于每个所述句子, 将该句子与其相似度值最高的所述图片放入候选负 样本对集; 将所述候选负样本对集中相似度值最高的预设数量的所述负样本对放 入所述负样本对集。

[0074] 该实施方式对应于跨文档训练目标 (cross-document objective) 中的确定正样 本对集和负样本对集, 此处不做具体赘述。

[0075] 在一个实施方式中, 所述训练文档为残次文档即该文档中部分句子和/或图 片被遮盖。

[0076] 该实施方式对应于次跨文档训练目标 (dropout sub-document objective), 此处 不做具体赘述。

[0077] 本说明书还提供的了一种模型训练装置, 应用于无监督图像文本匹配模型, 如上 面的实施方式所述。由于一种模型训练装置解决问题的原理与一种模型训练 方法相似, 因此一种模型训练装置的实施可以参见一种型训练方法的实施, 重复 之处不再赘述。以下所 使用的, 术语“单元”或者“模块”可以实现预定功能的 软件和/或硬件的组合。尽管以下实 施例所描述的装置较佳地以软件来实现, 但 是硬件, 或者软件和硬件的组合的实现也是可 能并被构想的。该装置具体可以包 括: 计算单元, 所述计算单元用于计算训练文档中图片 与句子的两两相似度值; 采样单元, 所述采样单元用于基于所述相似度值, 确定正样本对 集和负样本对集; 其中, 所述正样本对集中有预设数量的正样本对; 所述负样本对集中有 预设数量 的负样本对; 所述正样本对集和所述负样本对集用于进一步训练所述模型, 直至 预设数量的所述正样本对的平均相似度值大于预设数量的所述负样本对的平均 相似度 值, 且两者差值符合预设条件。

[0078] 本说明书还提供的了一种跨模态表征方法, 应用于无监督图像文本匹配模型 的训练或无监督图像文本匹配, 所述方法包括以下步骤。

[0079] 步骤S20: 根据获取的目标文档的图片信息, 获取对应图片的图片表征信息, 所述 图片表征信息包括该图片信息对应的至少一个对象, 以及概念信息; 所述概 念信息为根据 所述对象预测出的标签。

[0080] 步骤S22: 根据获取的所述目标文档的句子信息, 获取对应句子的句子表征 信息。

[0081] 步骤S24:将所述图片表征信息和所述句子表征信息进行匹配,得到对应于 该图片与该句子的匹配值;所述匹配值用于表征该图片与该句子的匹配度。

[0082] 在本实施方式中,所述对象可以用于表示图片信息中的物体。例如,请参 阅图5中的objects中的区域。所述对象的概念信息例如图5中concepts中的内 容信息。

[0083] 请参阅图5。在本实施方式中,由于引入了更多的同一文档内的图片句子对, 需要得到包含更细粒度信息的多模态表征,所以将图片分割为区域,对应不同的 对象,所述对象对应的概念信息以表征该图片;将句子分割为token,Transformer 可以看作是带有 attention机制的图网络,通过两个视觉/文本的Transformer对各模 态内的(对象/token)节点进行编码,与此同时引入了视觉的概念,即概念信息。将图片区域预测出的标签作为 图片包括的概念,将它们作为中间的桥梁将两个模 态的图桥接起来。概念会直接加入到视觉的图中,作为节点存在,而概念和文本 端的关系通过共享的embedding层来实现。这样的模型里,当句子里直接提到了 区域里对应的概念时,模型就能很快地捕捉到这样的匹配关系。

[0084] 在本实施方式中,所述图片信息对应的至少一个对象,以及所述对象对应的 概念信息以表征该图片。具体的,所述图片表征信息包括该图片信息对应的至少 一个对象,以及概念信息。为了在文档中提取可能的概念,从图像中检测到的视 觉实体,因为句子中的概念是模糊的,可能具有复合含义。使用一个预先训练的 Faster RCNN来提取出对象建议 $\{o_1, \dots, o_M\}$ 其中,每个对象 o_i 由其2048维感兴趣区域(ROI)特征 f_i 表示。同时,将这些对象的预测标签视为提取的概念 $\{c_1, \dots, c_M\}$ 。

[0085] 在本实施方式中,获取对应句子的句子表征信息即提取句子表示即所述句子 表征信息。具体的,一个句子首先被拆分为单词 $\{w_1, \dots, w_\lambda\}$ 然后将一个单词 w_j 和它的索引 i (w_i 在句子中的绝对位置)分别发送给一个300D的单词嵌入层和一个 位置嵌入层:

$$[0086] \quad \tilde{w}_i = \text{WordEmbed}(w_i)$$

$$[0087] \quad \tilde{u}_i = \text{PosEmbed}(i)$$

$$[0088] \quad \tilde{h}_i = \text{LayerNorm}(\tilde{w}_i + \tilde{u}_i) \quad (6)$$

[0089] 其中单词嵌入层使用基于GoogleNews预训练的词向量进行初始化,位置嵌 入层按照-0.02到0.02之间的均匀分布进行随机初始化。进一步项目的嵌入与前馈 层和编码由一个单模态NT层Transformer:

$$[0090] \quad h_i^0 = W_T \tilde{h}_i + b_T$$

$$[0091] \quad \{h_1^{l+1}, \dots, h_\lambda^{l+1}\} = \text{Transformer}_T^l(\{h_1^l, \dots, h_\lambda^l\}) \quad (7)$$

[0092] 其中, $W_T \in \mathbb{R}^{\text{dmulti} \times 300}$ 与 $b_T \in \mathbb{R}_{\text{dmulti}}$ 为前馈层的参数。

[0093] 最后,通过 $\{h_1^{\text{NT}}, \dots, h_\lambda^{\text{NT}}\}$ 上的平均池化层提取一个句子的表示。

[0094] 在本实施方式中,提取图像表示即所述图片表征信息。具体的,交叉模态 Transformer用于对图像中提取的可视对象和可视概念之间的依赖关系建模,对概 念和图像之间的链接建模。在这个交叉模态Transformer中,每个对象 o_j 用其ROI 特征 f_j 表示,并在其上添加一段嵌入,表示此片段为可视化对象:

$$[0095] \quad \tilde{f}_j = \text{LayerNorm}(W_V f_j + b_V)$$

$$[0096] \quad \tilde{s}_j = \text{LayerNorm}(\text{SegEmbed}(o_j))$$

$$[0097] \quad \tilde{v}_i = (\tilde{f}_i + \tilde{s}_i)/2 \quad (8)$$

[0098] 将每个概念 c_k 发送到单词嵌入层,通过前馈层投影,学习文本概念的嵌入,并在其上添加一段嵌入,表示此标记为可视化概念:

$$[0099] \quad \tilde{w}_k = \text{WordEmbed}(c_k)$$

$$[0100] \quad \tilde{t}_k = \text{LayerNorm}(W_T \tilde{w}_k + b_T)$$

$$[0101] \quad \tilde{s}_k = \text{LayerNorm}(\text{SegEmbed}(c_k))$$

$$[0102] \quad \tilde{c}_k = (\tilde{t}_k + \tilde{s}_k)/2 \quad (9)$$

[0103] 如果一个概念由几个单词组成,则应用平均池来获得文本嵌入。在提取句子表示时,单词嵌入层与层共享权重,如式6所示。当概念在句子中被直接提到时,概念和句子之间的联系就体现出来了。

[0104] 然后将每张图像中的所有对象和概念发送到NC层交叉模态Transformer:

$$[0105] \quad \{e_1^0, \dots, e_{2\mu}^0\} = \{\tilde{v}_1, \dots, \tilde{v}_\mu, \tilde{c}_1, \dots, \tilde{c}_\mu\}$$

$$[0106] \quad \{e_1^{l+1}, \dots, e_{2\mu}^{l+1}\} = \text{Transformer}_T^l(\{e_1^l, \dots, e_{2\mu}^l\}) \quad (10)$$

[0107] 类似地,输出的所有对象和概念上的平均池化层。在最后一个Transformer层 $\{e_1^{\text{NC}}, \dots, e_{2\mu}^{\text{NC}}\}$ 用于提取图像表示形式。

[0108] 在一个实施例中,对于一个文档 $d_i = \langle S_i, V_i \rangle$,由一组 $|S_i|$ 句子的 S_i 和一组 $|V_i|$ 图像的 V_i 组成,为了预测文档中每对图像和句子的标签,所有图片和句子生成特

[0109]

[0110] 性基于训练交叉模式表示学习模型和计算的相似矩阵 M_i ,其中第 (i, j) 个元素是第 i 个句子表示和第 j 个图像表示之间的余弦相似度。

[0111] 在MSCOCO (Lin et al.2014)和VIST (Huang et al.2016)上评估本说明书提出的方法。从这些众标数据集中收集图像和句子来组成文档,然后构造3个不同的数据集用于文档内链接预测:MSCOCO, DII和SIS。在MSCOCO中,每个文档由5组随机抽样的图像-标题对,5张干扰图像和5个干扰语句组成。在DII和SIS中,每个文档由来自同一个相册的5张图像和对应的独立描述(DII)或顺序故事(SIS)故事的5个句子组成。这些数据集的统计数据如表1所示。表1:数据集统计:密度 为文档ground-truth图中的边密度,即ground-truth图边的个数除以所有可能边的个数($n_i * m_i$)。

[0112] 表1:数据集统计

	train/val/test	n_i/m_i	# imgs	density
[0113] MSCOCO	25K/2K/2K	10/10	83K	5%
Story-DII	22K/3K/3K	5/5	47K	20%
Story-SIS	37K/5K/5K	5/5	76K	20%

[0114] 对于图像,在本场景示例中,使用由 (Anderson et al.2018) 提供的在视觉基因组 (Krishna et al.2017) 上预先训练的Faster-RCNN。在DII和SIS中,我们为每幅图像提取36个对象和概念,而数字在MSCOCO中是自适应的。设 $d_{\text{multi}}=1024, N_T=3, N_C=3$ 。每层中单模Transformer和交叉模Transformer有8头。Transformer中使用掩模来处理变长序列。对于训练目标,在函数TK中,设 $k=\min(n_i, m_i)$ 。子文档比例 p_{sub} 在SIS和DII中设置为0.6,在MSCOCO中设置为0.8。Margin α 在铰链损失设置为0.2。我们使用Adam optimizer训练本说明书提供的模型。在热身阶段,将学习速率从 $1e-7$ 线性提高到几个步骤后配置的最大学习速率。在MSCOCO中,最大学习率和热身步骤分别设置为 $1e-5$ 和3000,而在SIS和DII中设置为 $5e-5$ 和4000。在热身阶段之后,在经过3个学习周期的平台期后,学习速率都会降低5倍。小批尺寸为11。

[0115] 在无监督的多句子多图片文档内的跨模态链接预测任务上进行了实验,其中包括了基于MSCOCO,VIST构建出的三个文档数据集。对于每一个文档,其内部有多个句子和多个图片,需要去预测其中句子和图片间是否存在链接的边(匹配关系),使用AUC/P@1/P@5进行评估。相较于之前只使用cross-document objective的方法(表内MulLink),本说明书提供的方法有了明显的提高。

[0116] 表2:不同模型的综合性能。粗体数字表示每一列的最佳表现。

	MSCOCO		Story-DII		Story-SIS	
	AUC	p@1/p@5	AUC	p@1/p@5	AUC	p@1/p@5
[0117] Obj Detect	89.5	67.7/45.9	65.3	50.2/35.2	58.4	40.8/28.6
NoStruct	87.4	50.6/34.3	77.0	60.8/46.3	64.5	42.8/33.2
MulLink	99.0	95.0/81.1	82.9	72.0/55.8	68.8	51.8/38.6
Ours	99.3	97.6/86.0	85.5	77.2/60.1	70.2	53.1/39.8

[0118] 同时在本场景示例中,对本说明书提出的模型的结构,和三个部分的训练目标进行了消融实验,结果如表3:

[0119] 表3.部分消融实验的结果

	backbone	S	AUC	p@1/p@5
[0120]	1 Transformer	T	85.5	77.2/60.1
	2 Transformer w/o Concept	T	85.3	75.8/59.8
	3 w/o Transformer	T	85.1	75.0/59.0
	4 w/o Transformer&Concept	T	85.1	74.6/59.1
	5 GRU+CNN	T	84.0	72.9/58.0
	6 Transformer	O	85.0	75.5/59.4
	7 Transformer w/o Concept	O	84.8	73.4/58.8
	8 w/o Transformer	O	84.4	73.1/58.2
	9 GRU+CNN	O	82.8	72.0/55.8

[0121] 其中,S列代表采样方法,T代表同时使用三种目标训练,O代表只使用跨文档训练目标,w/o代表without,w/o Transformer的方法里我们使用GRU对句子进行表征,对图片的各个区域进行softmax pooling进行表征。

[0122] 可以看到整体上同时使用三种目标可以采样到更多的信息,帮助训练。同时本说

说明书提出的模型更好地利用了细粒度的信息,也获得了更好地跨模态表征。

[0123] 同时,和有监督、无监督和迁移学习的比较。有监督的方法直接使用文档内的匹配的图片句子对作为训练,如图6,迁移学习则尝试迁移从MSCOCO上进行有监督训练的信息到DII测试集上,如表4。

[0124] 表4.迁移学习和无监督学习的比较

Method	AUC	p@1/p@5
1 Transfer from MSCOCO	78.6	66.5/49.5
2 Unsupervised	85.5	77.2/60.1

[0126] 可以看到相较于只使用跨文档训练目标,同时使用三种目标得到的更多样本 对里包括了更多的信息,本说明书提供的无监督的方法可以利用训练集内更多的 信息(~40%),相较于迁移自其他数据集的信息,也更加有效。

[0127] 在一个实施方式中,所述概念信息和所述句子信息的表征信息共享嵌入层。

[0128] 本说明书还提供了一种无监督图像文本匹配方法,所述方法可以包括以下步骤。

[0129] 步骤S30:基于上述的模型训练方法,训练无监督图像文本匹配模型。

[0130] 步骤S32:基于所述无监督图像文本匹配模型,获取目标文档的图片信息的 图片表征信息,所述图片表征信息包括该图片信息对应的至少一个对象,以及概念信息;所述概念信息为根据所述对象预测出的标签。

[0131] 步骤S34:基于所述无监督图像文本匹配模型,获取所述目标文档对应句子的 句子表征信息。

[0132] 步骤S36:将所述图片表征信息和所述句子表征信息进行匹配,得到对应于 该图片与该句子的匹配值;所述匹配值用于表征该图片与该句子的匹配度。

[0133] 上述步骤的先后顺序不做具体限定。

[0134] 在本实施方式中,仅针对与前述实施方式不同之处进行了描述,其它内容可 以对照前述实施方式的内容解释,在此不再赘述。

[0135] 本说明书还提供了一种无监督图像文本匹配装置,所述装置包括:

[0136] 训练单元,所述训练单元用于接收跨文档训练、文档内部的训练、次跨文档 训练的训练样本,以进行训练;

[0137] 无监督图像文本匹配单元,所述无监督图像文本匹配单元用于基于所述训练 单元,匹配目标文档的中的图片与句子;所述无监督图像文本匹配单元至少包括:

[0138] 第一提取单元,所述第一提取单元用于获取所述目标文档的图片对应的图片 表征信息,所述图片表征信息包括该图片信息对应的至少一个对象,以及概念信 息;所述概念信息为根据所述对象预测出的标签;

[0139] 第二提取单元,所述第二提取单元用于获取所述目标文档对应句子的句子表 征信息;

[0140] 匹配单元,所述匹配单元用于将所述图片表征信息和所述句子表征信息进行 匹配,得到对应于该图片与该句子的匹配值;所述匹配值用于表征该图片与该句 子的匹配度。

[0141] 在本实施方式中,仅针对与前述实施方式不同之处进行了描述,其它内容可 以对照前述实施方式的内容解释,在此不再赘述。

[0142] 上述实施例阐明的装置或模块等,具体可以由计算机芯片或实体实现,或者由具有某种功能的产品来实现。为了描述的方便,描述以上装置时以功能分为各种模块分别描述。当然,在实施本申请时可以把各模块的功能在同一个或多个软件和/或硬件中实现,也可以将实现同一功能的模块由多个子模块的组合实现等。以上所描述的装置实施例仅仅是示意性的,例如,所述模块的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式,例如多个模块或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。

[0143] 本领域技术人员也知道,除了以纯计算机可读程序代码方式实现控制器以外,完全可以通过将方法步骤进行逻辑编程来使得控制器以逻辑门、开关、专用集成电路、可编程逻辑控制器和嵌入微控制器等的形式来实现相同功能。因此这种控制器可以被认为是一种硬件部件,而对其内部包括的用于实现各种功能的装置也可以视为硬件部件内的结构。或者甚至,可以将用于实现各种功能的装置视为既可以是实现方法的软件模块又可以是硬件部件内的结构。

[0144] 本申请可以在由计算机执行的计算机可执行指令的一般上下文中描述,例如程序模块。一般地,程序模块包括执行特定任务或实现特定抽象数据类型的例程、程序、对象、组件、数据结构、类等等。也可以在分布式计算环境中实践本申请,在这些分布式计算环境中,由通过通信网络而被连接的远程处理设备来执行任务。在分布式计算环境中,程序模块可以位于包括存储设备在内的本地和远程计算机存储介质中。

[0145] 通过以上的实施方式的描述可知,本领域的技术人员可以清楚地了解到本申请可借助软件加必需的通用硬件平台的方式来实现。基于这样的理解,本申请的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来,该计算机软件产品可以存储在存储介质中,如ROM/RAM、磁碟、光盘等,包括若干指令用以使得一台计算机设备(可以是个人计算机,移动终端,服务器,或者网络设备等)执行本申请各个实施例或者实施例的某些部分所述的方法。

[0146] 本说明书中的各个实施例采用递进的方式描述,各个实施例之间相同或相似的部分互相参见即可,每个实施例重点说明的都是与其他实施例的不同之处。本申请可用于众多通用或专用的计算机系统环境或配置中。例如:个人计算机、服务器计算机、手持设备或便携式设备、平板型设备、多处理器系统、基于微处理器的系统、置顶盒、可编程的电子设备、网络PC、小型计算机、大型计算机、包括以上任何系统或设备的分布式计算环境等等。

[0147] 虽然通过实施例描绘了本申请,本领域普通技术人员知道,本申请有许多变形和变化而不脱离本申请的精神,希望所附的权利要求包括这些变形和变化而不脱离本申请。

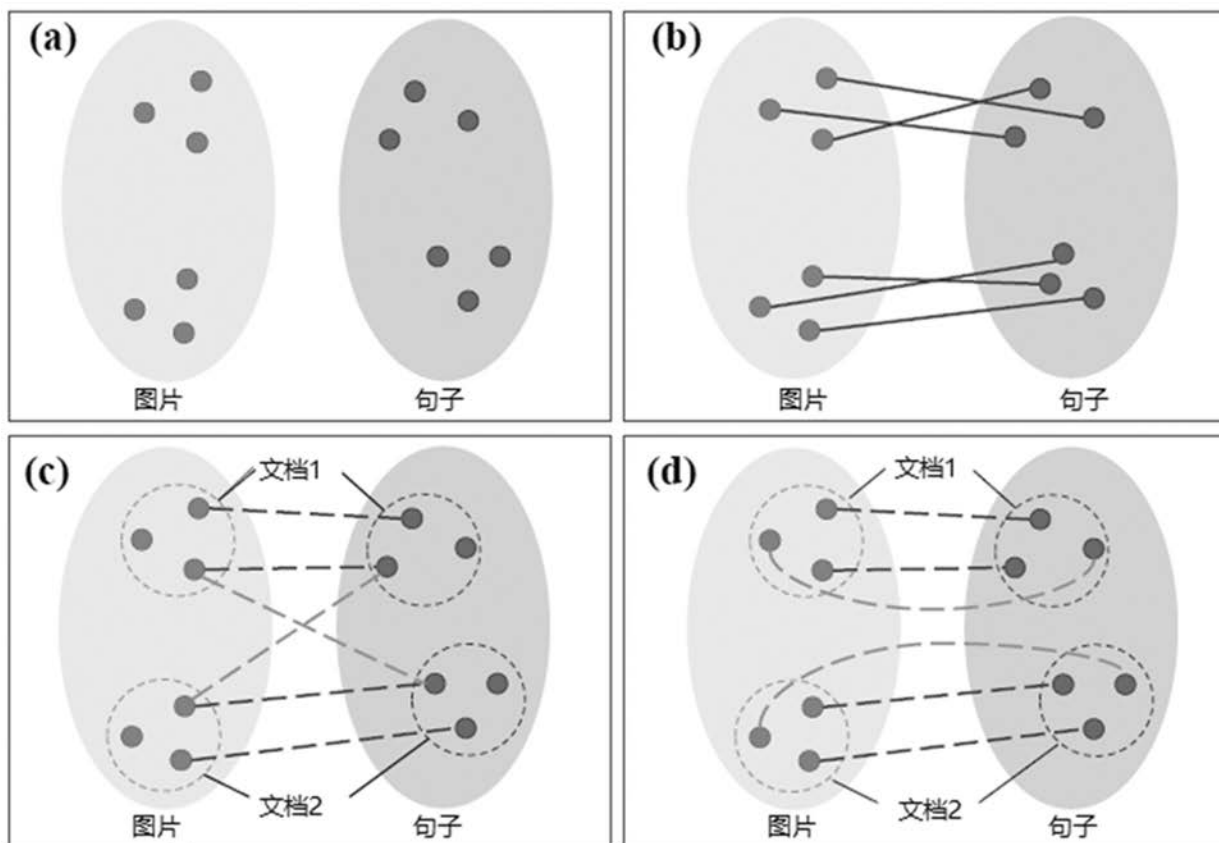


图1

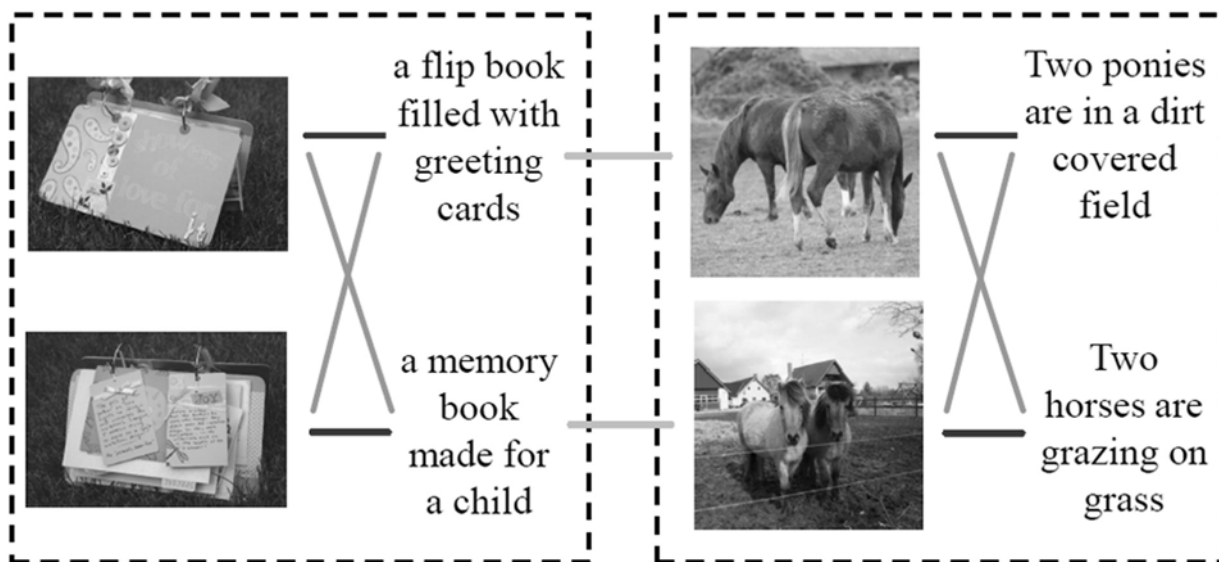


图2

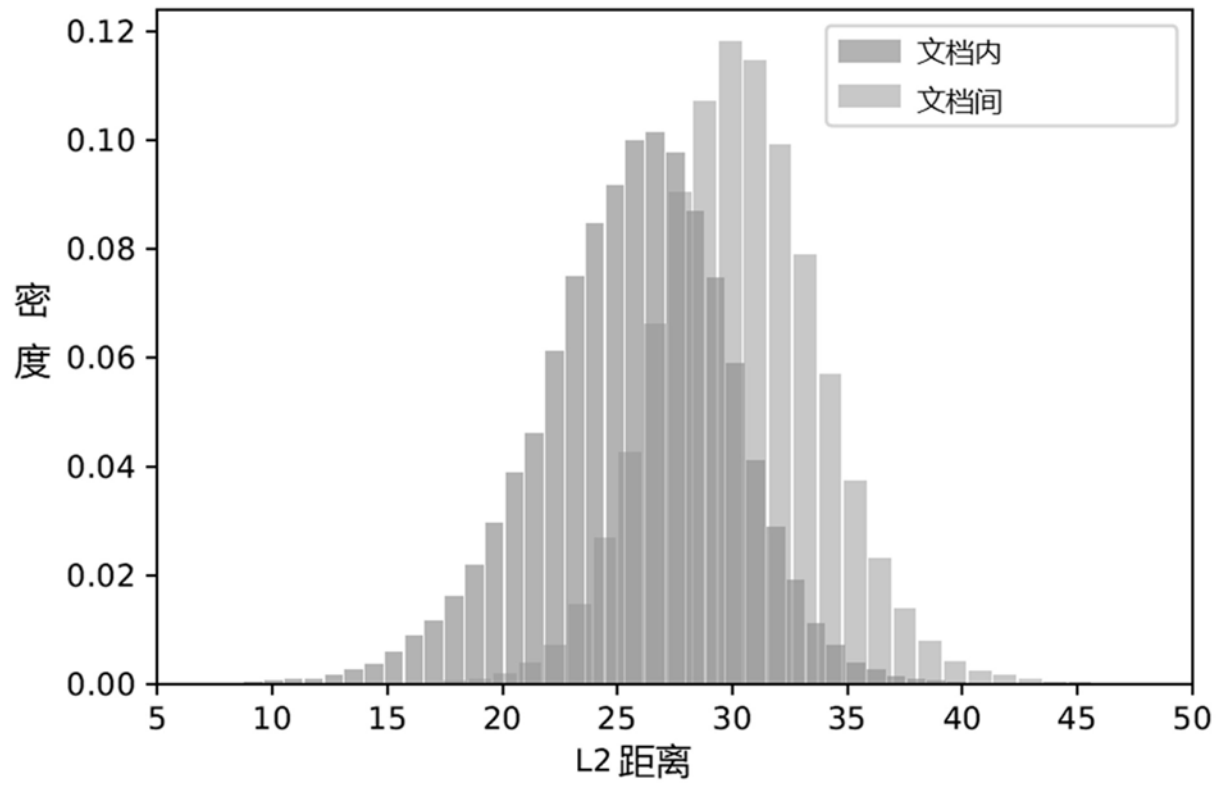


图3

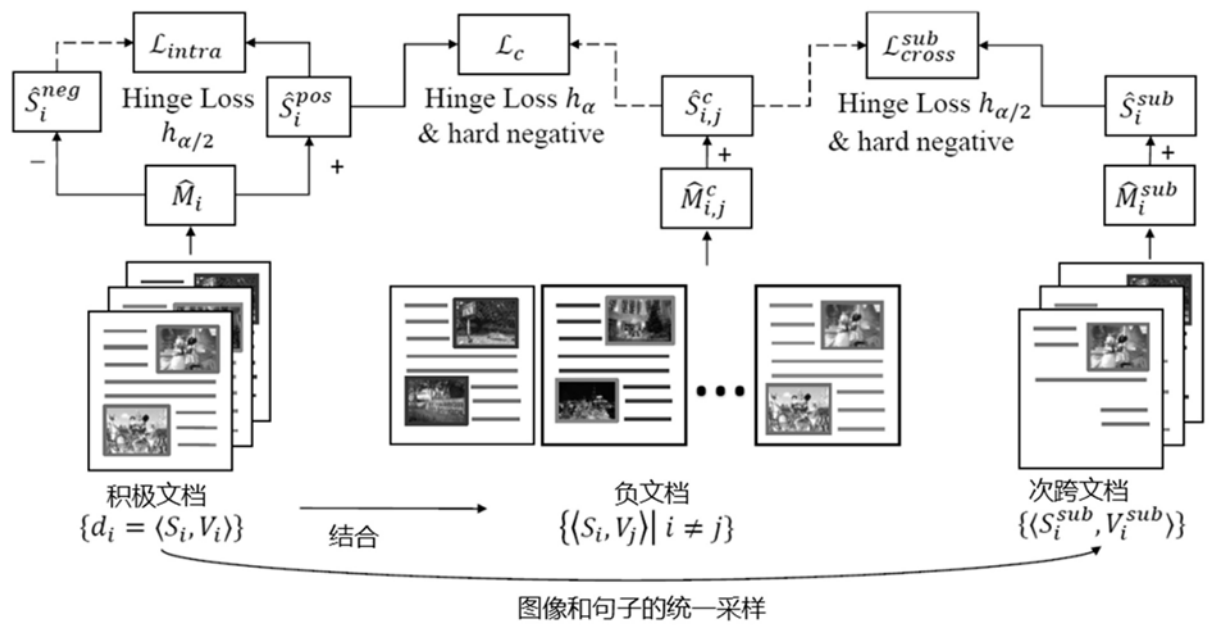


图4

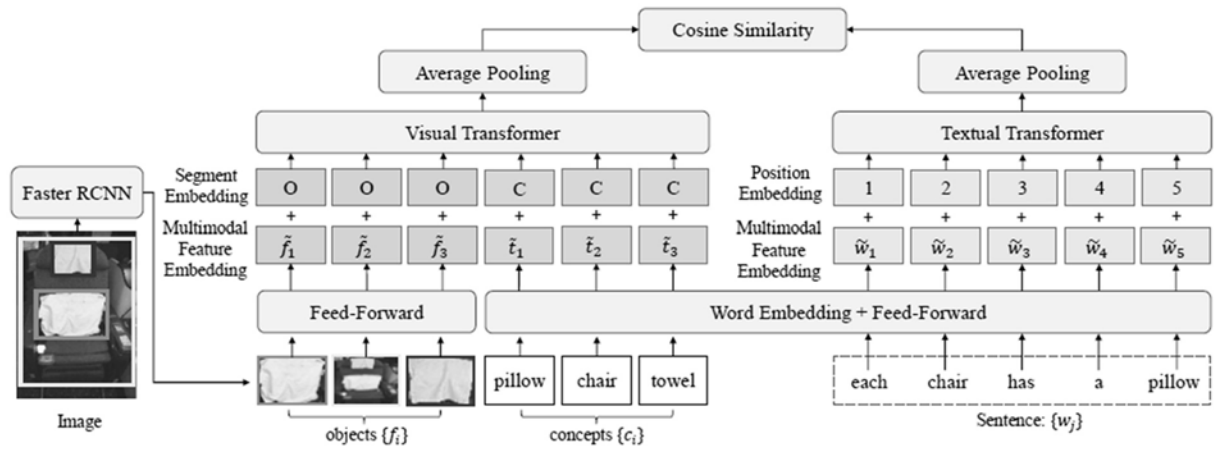


图5

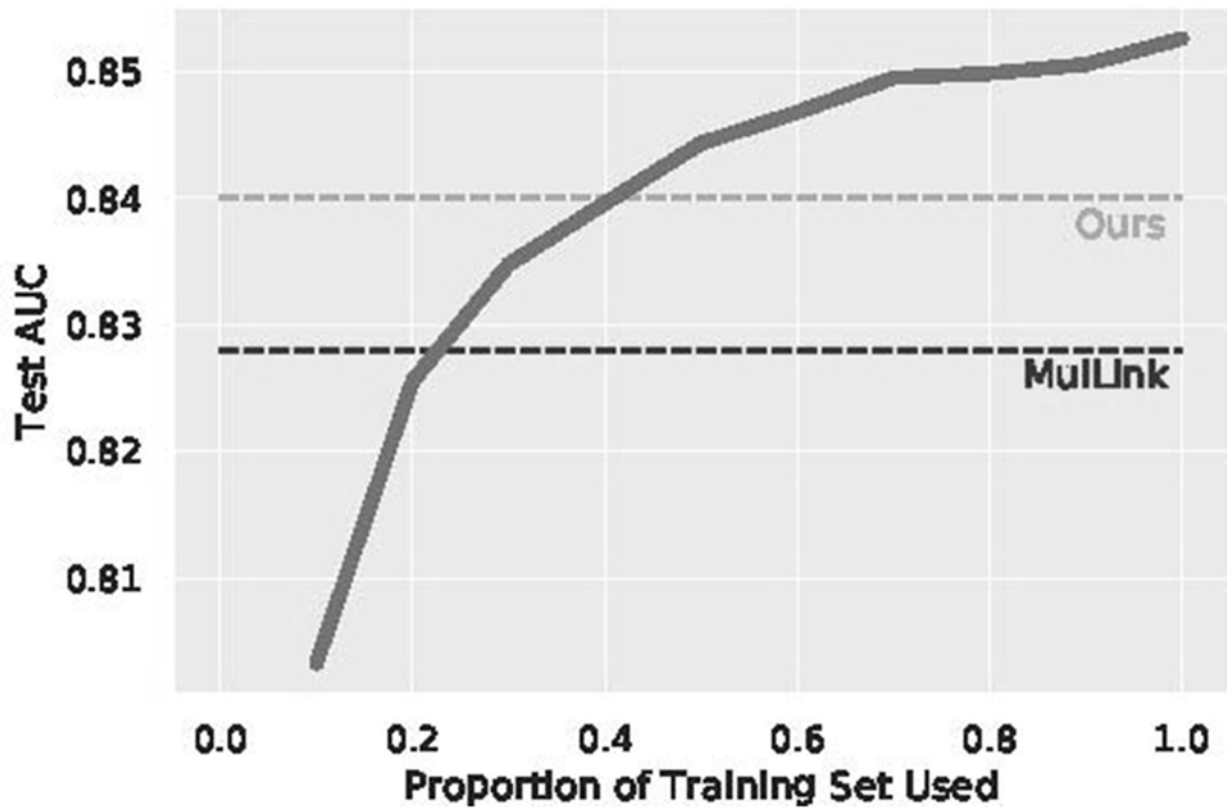


图6