



(12)发明专利申请

(10)申请公布号 CN 108960959 A  
(43)申请公布日 2018.12.07

(21)申请号 201810501840.4

(22)申请日 2018.05.23

(71)申请人 山东大学

地址 250101 山东省济南市高新区舜华路  
中段

(72)发明人 刘金环 宋雪萌 马军 甘甜  
聂礼强

(74)专利代理机构 济南圣达知识产权代理有限公司 37221

代理人 黄海丽

(51)Int.Cl.

G06Q 30/06(2012.01)

G06N 3/02(2006.01)

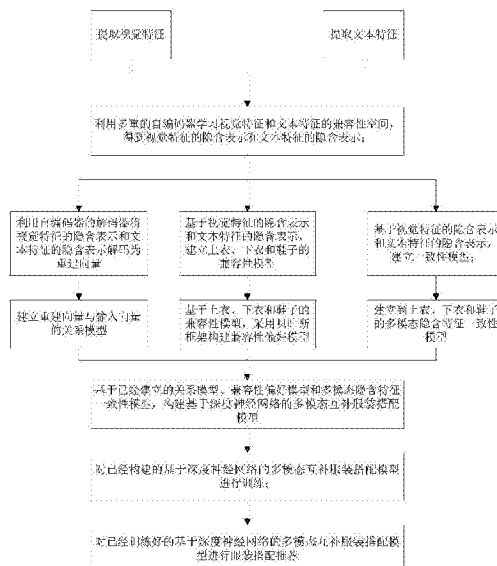
权利要求书3页 说明书8页 附图3页

(54)发明名称

基于神经网络的多模态互补服装搭配方法、系统  
及介质

(57)摘要

本发明公开了基于神经网络的多模态互补服装搭配方法、系统及介质,从服装的图片中获取视觉特征,从服装的文字描述中获取文本特征;利用自编码器学习不同服装的视觉特征和文本特征的兼容性空间,得到视觉特征的隐含表示和文本特征的隐含表示;建立重建向量分别与输入特征之间的关系模型;建立服装兼容性模型;然后,基于服装兼容性模型,采用贝叶斯个性化排序算法构建兼容性偏好模型;建立视觉特征隐含表示和文本特征隐含表示的一致性模型;然后,建立服装的多模态隐含特征一致性模型;构建基于深度神经网络的多模态互补服装搭配模型;对已经构建的多模态互补服装搭配模型进行训练;利用已经训练好的多模态互补服装搭配模型进行服装搭配推荐。



CN 108960959 A

1. 基于神经网络的多模态互补服装搭配方法,其特征是,包括:

步骤(1):从服装的图片中获取视觉特征,同时,从服装的文字描述中获取文本特征;

步骤(2):利用自编码器学习不同服装的视觉特征和文本特征的兼容性空间,得到视觉特征的隐含表示和文本特征的隐含表示;

步骤(3):利用解码器将步骤(2)得到的视觉特征的隐含表示和文本特征的隐含表示解码为重建向量;建立重建向量分别与输入特征之间的关系模型;

步骤(4):基于步骤(2)得到的视觉特征的隐含表示和文本特征的隐含表示,建立服装兼容性模型;然后,基于服装兼容性模型,采用贝叶斯个性化排序算法构建兼容性偏好模型;

步骤(5):基于步骤(2)得到的视觉特征的隐含表示和文本特征的隐含表示,建立视觉特征隐含表示和文本特征隐含表示的一致性模型;然后,建立服装的多模态隐含特征一致性模型;

步骤(6):基于步骤(3)、步骤(4)和步骤(5)的计算结果,构建基于神经网络的多模态互补服装搭配模型;对已经构建的多模态互补服装搭配模型进行训练;利用已经训练好的多模态互补服装搭配模型进行服装搭配推荐。

2. 如权利要求1所述的基于神经网络的多模态互补服装搭配方法,其特征是,步骤(1)中:

所述从服装的图片中获取视觉特征,是通过深度卷积神经网络从服装的图片中获取视觉特征;

所述从服装的文字描述中获取文本特征,是通过词袋模型从服装的文字描述中获取文本特征。

3. 如权利要求1所述的基于神经网络的多模态互补服装搭配方法,其特征是,步骤(2)中:

将上衣 $t_i$ 的视觉特征 $v_i^t$ 、上衣 $t_i$ 的文本特征 $c_i^t$ 、下衣 $b_j$ 的视觉特征 $v_j^b$ 、下衣 $b_j$ 的文本特征 $c_j^b$ 、鞋子 $s_k$ 的视觉特征 $v_k^s$ 和鞋子 $s_k$ 的文本特征 $c_k^s$ 输入到自编码器的编码器中;输出上衣 $t_i$ 的视觉隐含表示 $\tilde{v}_i^t$ 、上衣 $t_i$ 的文本隐含表示 $\tilde{c}_i^t$ 、下衣 $b_j$ 的视觉隐含表示 $\tilde{v}_j^b$ 、下衣的文本隐含表示 $\tilde{c}_j^b$ 、鞋子 $s_k$ 的视觉隐含表示 $\tilde{v}_k^s$ 和鞋子 $s_k$ 的文本隐含表示 $\tilde{c}_k^s$ 。

4. 如权利要求3所述的基于神经网络的多模态互补服装搭配方法,其特征是,步骤(3)中:

通过自编码器的解码器,将上衣 $t_i$ 的视觉隐含表示 $\tilde{v}_i^t$ 、上衣 $t_i$ 的文本隐含表示 $\tilde{c}_i^t$ 、下衣 $b_j$ 的视觉隐含表示 $\tilde{v}_j^b$ 、下衣的文本隐含表示 $\tilde{c}_j^b$ 、鞋子 $s_k$ 的视觉隐含表示 $\tilde{v}_k^s$ 、鞋子 $s_k$ 的文本隐含表示 $\tilde{c}_k^s$ 解码为上衣 $t_i$ 的视觉重建向量 $\hat{v}_i^t$ 、上衣 $t_i$ 的文本重建向量 $\hat{c}_i^t$ 、下衣 $b_j$ 的重建向量 $\hat{v}_j^b$ 、下衣的重建向量 $\hat{c}_j^b$ 、鞋子 $s_k$ 的视觉重建向量 $\hat{v}_k^s$ 、鞋子 $s_k$ 的文本重建向量 $\hat{c}_k^s$ 。

5. 如权利要求1所述的基于神经网络的多模态互补服装搭配方法,其特征是,步骤(3)中:建立重建向量与输入特征之间的关系模型的步骤为:

$$l_{AE}(x) = l(v_x) + l(c_x)$$

$$= \frac{1}{2} \|\hat{v}_x - v_x\|_2^2 + \frac{1}{2} \|\hat{c}_x - c_x\|_2^2;$$

其中,  $l_{AE}(x)$  表示服装 $x$ 重建向量与视觉特征和文本特征之间的关系模型;  $l(v_x)$  表示服装 $x$ 的视觉重建向量与视觉特征的重建误差;  $l(c_x)$  表示服装 $x$ 的文本重建向量与文本特征的重建误差;  $\hat{v}_x$  表示服装 $x$ 的视觉重建向量;  $v_x$  表示服装 $x$ 的视觉特征;  $\hat{c}_x$  表示服装 $x$ 的文本重建向量;  $c_x$  表示服装 $x$ 的文本特征;

建立整个服装的重建向量与输入特征之间的关系模型:

$$l_{AE} = l_{AE}(t_i) + l_{AE}(b_j) + l_{AE}(s_k)。$$

6. 如权利要求1所述的基于神经网络的多模态互补服装搭配方法,其特征是,步骤(4)中:建立服装兼容性模型 $comp_{ijk}$ :

$$comp_{ijk} = comp_{ik} + comp_{jk};$$

$$comp_{ik} = (1 - \beta)(\tilde{v}_i^t)^T \tilde{v}_k^s + \beta(\tilde{v}_i^t)^T \tilde{v}_k^s;$$

$$comp_{jk} = (1 - \beta)(\tilde{v}_j^b)^T \tilde{v}_k^s + \beta(\tilde{v}_j^b)^T \tilde{v}_k^s;$$

其中,  $comp_{ik}$  表示上衣和鞋子的兼容性模型;  $comp_{jk}$  表示下衣和鞋子的兼容性模型;

基于服装兼容性模型,采用贝叶斯个性化排序算法构建兼容性偏好模型 $l_{bpr}$ :

$$l_{bpr} = \sum_{(i,j,k_+,k_-)} -\ln(s(comp_{ijk_+} - comp_{ijk_-}));$$

其中,  $comp_{ijk_+}$  表示上衣和下衣对正例鞋子的兼容性偏好,  $comp_{ijk_-}$  表示上衣和下衣对负例鞋子的兼容性偏好;  $s(comp_{ijk_+} - comp_{ijk_-})$  表示神经网络的阈值函数;  $(i, j, k_+, k_-)$  表示上衣 $t_i$ 和下衣 $b_j$ 相对于鞋子 $s_{k_-}$ 来说更适合与鞋子 $s_{k_+}$ 搭配。

7. 如权利要求1所述的基于神经网络的多模态互补服装搭配方法,其特征是,步骤(5)中:视觉特征隐含表示和文本特征隐含表示的一致性模型 $l_{vc}(t_i)$ :

$$l_{vc}(t_i) = -\ln(s((\tilde{v}_i^t)^T \tilde{c}_i^t));$$

其中,  $s((\tilde{v}_i^t)^T \tilde{c}_i^t)$  表示神经网络的阈值函数;  $\tilde{v}_i^t$  表示服装 $t_i$ 的视觉特征隐含表示;  $\tilde{c}_i^t$  表示服装 $t_i$ 的文本特征隐含表示;

服装的多模态隐含特征一致性模型 $l_{mod}$ :

$$l_{mod} = \sum_{(i,j,k_+,k_-)} l_{vc}(t_i) + l_{vc}(b_j) + l_{vc}(s_{k_+}) + l_{vc}(s_{k_-});$$

其中,  $l_{vc}(t_i)$  表示上衣视觉特征隐含表示和文本特征隐含表示的一致性模型;  $l_{vc}(b_j)$  表示下衣视觉特征隐含表示和文本特征隐含表示的一致性模型;  $l_{vc}(s_{k_+})$  表示正例鞋子视觉特征隐含表示和文本特征隐含表示的一致性模型;  $l_{vc}(s_{k_-})$  表示负例鞋子视觉特征隐含表示和文本特征隐含表示的一致性模型。

8. 如权利要求1所述的基于神经网络的多模态互补服装搭配方法,其特征是,步骤(6)中:基于深度神经网络的多模态互补服装搭配模型:

$$L = l_{AE} + l_{mod} + l_{bpr};$$

对已经构建的多模态互补服装搭配模型参数通过随机梯度下降方法进行训练,通过迭

代至该模型收敛,输出最终参数;

利用已经训练好的基于神经网络的多模态互补服装搭配进行服装搭配推荐;

通过训练好的最终参数,计算所有 $comp_{ijk}$ 的值;

其中 $comp_{ijk}$ 最大的值所对应的鞋子即为上衣和下衣最为搭配的鞋子。

9. 基于神经网络的多模态互补服装搭配系统,其特征是,包括:存储器、处理器以及存储在存储器上并在处理器上运行的计算机指令,所述计算机指令被处理器运行时,完成上述权利要求1-8任一方法所述的步骤。

10. 一种计算机可读存储介质,其特征是,其上存储有计算机指令,所述计算机指令被处理器运行时,完成上述权利要求1-8任一方法所述的步骤。

## 基于神经网络的多模态互补服装搭配方法、系统及介质

### 技术领域

[0001] 本发明涉及基于神经网络的多模态互补服装搭配方法、系统及介质。

### 背景技术

[0002] 现如今,除了对服装的生理需求外,越来越多的人开始注重和追寻穿着的时尚、优雅、大方得体等等。然而,并不是每个人对服装搭配都有很好的品味。当面对海量的服装商品时,很多人会感到服装搭配非常的困难和厌烦。因此,我们开发一套有效的服装搭配方案来帮助人们为给定的服装找出协调又时尚的搭配。当前的服装搭配技术主要包括基于协同过滤的方法和基于内容的方法。其中,前者通过具有相似品味和偏好的用户历史行为进行推荐,如:用户的购买行为,用户对商品的文本描述和其它用户的点击购买行为等。这种方法存在冷启动的问题,即无法为一个没有任何相关历史行为的物品或用户进行推荐。后者基于物品间的视觉兼容性进行推荐。这种方法通常只考虑物品的视觉信息,不能全面的对物品之间的兼容性进行建模。另外,在服装搭配过程中还存在着稀疏性的问题。

### 发明内容

[0003] 为了解决现有技术的不足,本发明提供了基于神经网络的多模态互补服装搭配方法、系统及介质,可以有效解决服装之间的稀疏性问题,并且可以通过挖掘物品的多模态关系全面的对不同服装间的兼容性进行建模。

[0004] 作为本发明的第一方面,提供了基于神经网络的多模态互补服装搭配方法;

[0005] 基于神经网络的多模态互补服装搭配方法,包括:

[0006] 步骤(1):从服装的图片中获取视觉特征,同时,从服装的文字描述中获取文本特征;

[0007] 步骤(2):利用自编码器学习不同服装的视觉特征和文本特征的兼容性空间,得到视觉特征的隐含表示和文本特征的隐含表示;

[0008] 步骤(3):利用多重解码器将步骤(2)得到的视觉特征的隐含表示和文本特征的隐含表示解码为重建向量;建立重建向量分别与输入特征之间的关系模型;

[0009] 步骤(4):基于步骤(2)得到的视觉特征的隐含表示和文本特征的隐含表示,建立服装兼容性模型;然后,基于服装兼容性模型,采用贝叶斯个性化排序算法构建兼容性偏好模型;

[0010] 步骤(5):基于步骤(2)得到的视觉特征的隐含表示和文本特征的隐含表示,建立视觉特征隐含表示和文本特征隐含表示的一致性模型;然后,建立服装的多模态隐含特征一致性模型;

[0011] 步骤(6):基于步骤(3)、步骤(4)和步骤(5)的计算结果,构建基于深度神经网络的多模态互补服装搭配模型;对已经构建的多模态互补服装搭配模型进行训练;利用已经训练好的多模态互补服装搭配模型进行服装搭配推荐。

[0012] 作为本发明的进一步改进,步骤(1)中:

- [0013] 所述服装,包括:上衣、下衣和鞋子;
- [0014] 所述服装的图片,是指上衣、下衣或鞋子的彩色图片;
- [0015] 所述服装的文字描述,包括:服装的样式、功能和类别;
- [0016] 所述视觉特征,包括:上衣的视觉特征、下衣的视觉特征和鞋子的视觉特征;
- [0017] 所述文本特征,包括:上衣的文本特征、下衣的文本特征和鞋子的文本特征;
- [0018] 作为本发明的进一步改进,步骤(1)中:
- [0019] 所述从服装的图片中获取视觉特征,是通过深度卷积神经网络从服装的图片中获取视觉特征;
- [0020] 所述从服装的文字描述中获取文本特征,是通过词袋模型从服装的文字描述中获取文本特征。
- [0021] 作为本发明的进一步改进,步骤(2)中:
- [0022] 将上衣 $t_i$ 的视觉特征 $v_i^t$ 、上衣 $t_i$ 的文本特征 $c_i^t$ 、下衣 $b_j$ 的视觉特征 $v_j^b$ 、下衣 $b_j$ 的文本特征 $c_j^b$ 、鞋子 $s_k$ 的视觉特征 $v_k^s$ 和鞋子 $s_k$ 的文本特征 $c_k^s$ 输入到自编码器的编码器中;输出上衣 $t_i$ 的视觉隐含表示 $\hat{v}_i^t$ 、上衣 $t_i$ 的文本隐含表示 $\hat{c}_i^t$ 、下衣 $b_j$ 的视觉隐含表示 $\hat{v}_j^b$ 、下衣的文本隐含表示 $\hat{c}_j^b$ 、鞋子 $s_k$ 的视觉隐含表示 $\hat{v}_k^s$ 和鞋子 $s_k$ 的文本隐含表示 $\hat{c}_k^s$ ;
- [0023] 作为本发明的进一步改进,步骤(3)中:
- [0024] 通过自编码器的解码器,将上衣 $t_i$ 的视觉隐含表示 $\hat{v}_i^t$ 、上衣 $t_i$ 的文本隐含表示 $\hat{c}_i^t$ 、下衣 $b_j$ 的视觉隐含表示 $\hat{v}_j^b$ 、下衣的文本隐含表示 $\hat{c}_j^b$ 、鞋子 $s_k$ 的视觉隐含表示 $\hat{v}_k^s$ 、鞋子 $s_k$ 的文本隐含表示 $\hat{c}_k^s$ 解码为上衣 $t_i$ 的视觉重建向量 $\hat{v}_i^t$ 、上衣 $t_i$ 的文本重建向量 $\hat{c}_i^t$ 、下衣 $b_j$ 的重建向量 $\hat{v}_j^b$ 、下衣的重建向量 $\hat{c}_j^b$ 、鞋子 $s_k$ 的视觉重建向量 $\hat{v}_k^s$ 、鞋子 $s_k$ 的文本重建向量 $\hat{c}_k^s$ ;
- [0025] 作为本发明的进一步改进,步骤(3)中:建立重建向量与输入特征之间的关系模型的步骤为:
- [0026] 
$$l_{AE}(x) = l(v_x) + l(c_x)$$
- [0027] 
$$= \frac{1}{2} \|\hat{v}_x - v_x\|_2^2 + \frac{1}{2} \|\hat{c}_x - c_x\|_2^2;$$
- [0028] 其中, $l_{AE}(x)$ 表示服装 $x$ 重建向量与视觉特征和文本特征之间的关系模型; $l(v_x)$ 表示服装 $x$ 的视觉重建向量与视觉特征的重建误差; $l(c_x)$ 表示服装 $x$ 的文本重建向量与文本特征的重建误差; $\hat{v}_x$ 表示服装 $x$ 的视觉重建向量; $v_x$ 表示服装 $x$ 的视觉特征; $\hat{c}_x$ 表示服装 $x$ 的文本重建向量; $c_x$ 表示服装 $x$ 的文本特征;
- [0029] 建立整个服装的重建向量与输入特征之间的关系模型:
- [0030]  $l_{AE} = l_{AE}(t_i) + l_{AE}(b_j) + l_{AE}(s_k)$ 。
- [0031] 作为本发明的进一步改进,步骤(4)中:建立服装兼容性模型 $comp_{ijk}$ :
- [0032]  $comp_{ijk} = comp_{ik} + comp_{jk}$ ;
- [0033]  $comp_{ik} = (1 - \beta)(\hat{v}_i^t)^T \hat{v}_k^s + \beta(\hat{v}_i^t)^T \hat{v}_k^s$ ;
- [0034]  $comp_{jk} = (1 - \beta)(\hat{v}_j^b)^T \hat{v}_k^s + \beta(\hat{v}_j^b)^T \hat{v}_k^s$ 。
- [0035] 其中, $comp_{ik}$ 表示上衣和鞋子的兼容性模型; $comp_{jk}$ 表示下衣和鞋子的兼容性模

型；

[0036] 作为本发明的进一步改进，步骤(4)中：基于服装兼容性模型，采用贝叶斯个性化排序算法构建兼容性偏好模型 $l_{bpr}$ ：

$$[0037] \quad l_{bpr} = \sum_{(i,j,k_+,k_-)} -\ln \left( s(\text{comp}_{ijk_+} - \text{comp}_{ijk_-}) \right);$$

[0038] 其中， $\text{comp}_{ijk_+}$ 表示上衣和下衣对正例鞋子的兼容性偏好， $\text{comp}_{ijk_-}$ 表示上衣和下衣对负例鞋子的兼容性偏好； $s(\text{comp}_{ijk_+} - \text{comp}_{ijk_-})$ 表示神经网络的阈值函数；(i, j,  $k_+$ ,  $k_-$ )表示上衣 $t_i$ 和下衣 $b_j$ 相对于鞋子 $s_{k_-}$ 来说更适合与鞋子 $s_{k_+}$ 搭配。

[0039] 作为本发明的进一步改进，步骤(5)中：视觉特征隐含表示和文本特征隐含表示的一致性模型 $l_{vc}(t_i)$ ：

$$[0040] \quad l_{vc}(t_i) = -\ln \left( s((\tilde{v}_i^t)^T \tilde{c}_i^t) \right);$$

[0041] 其中， $s(\tilde{v}_i^t \tilde{c}_i^t)$ 表示神经网络的阈值函数； $\tilde{v}_i^t$ 表示服装 $t_i$ 的视觉特征隐含表示； $\tilde{c}_i^t$ 表示服装 $t_i$ 的文本特征隐含表示。

[0042] 作为本发明的进一步改进，步骤(5)中：服装的多模态隐含特征一致性模型 $l_{mod}$ ：

$$[0043] \quad l_{mod} = \sum_{(i,j,k_+,k_-)} l_{vc}(t_i) + l_{vc}(b_j) + l_{vc}(s_{k_+}) + l_{vc}(s_{k_-});$$

[0044] 其中， $l_{vc}(t_i)$ 表示上衣视觉特征隐含表示和文本特征隐含表示的一致性模型； $l_{vc}(b_j)$ 表示下衣视觉特征隐含表示和文本特征隐含表示的一致性模型； $l_{vc}(s_{k_+})$ 表示正例鞋子视觉特征隐含表示和文本特征隐含表示的一致性模型； $l_{vc}(s_{k_-})$ 表示负例鞋子视觉特征隐含表示和文本特征隐含表示的一致性模型。

[0045] 作为本发明的进一步改进，步骤(6)中：基于深度神经网络的多模态互补服装搭配模型：

$$[0046] \quad L = l_{AE} + l_{mod} + l_{bpr};$$

[0047] 作为本发明的进一步改进，步骤(6)中：对已经构建的多模态互补服装搭配模型参数通过随机梯度下降方法进行训练，通过迭代至该模型收敛，输出最终参数。

[0048] 作为本发明的进一步改进，步骤(6)中：利用已经训练好的基于深度神经网络的多模态互补服装搭配进行服装搭配推荐：

[0049] 通过训练好的最终参数，计算所有 $\text{comp}_{ijk}$ 的值；

[0050] 其中 $\text{comp}_{ijk}$ 最大的值所对应的鞋子即为上衣和下衣最为搭配的鞋子。

[0051] 作为本发明的第二方面，提供了基于神经网络的多模态互补服装搭配系统；

[0052] 基于神经网络的多模态互补服装搭配系统，包括：存储器、处理器以及存储在存储器上并在处理器上运行的计算机指令，所述计算机指令被处理器运行时，完成上述任一方法所述的步骤。

[0053] 作为本发明的第三方面，提供了一种计算机可读存储介质；

[0054] 一种计算机可读存储介质，其上存储有计算机指令，所述计算机指令被处理器运行时，完成上述任一方法所述的步骤。

[0055] 与现有技术相比，本发明的有益效果是：

[0056] 1、基于深度神经网络的多模态互补服装搭配方法可以对多个互补商品进行搭配；

- [0057] 2、该方法可以无缝的挖掘商品间的多模态信息(即视觉和文本模态)；
- [0058] 3、该模型可有效解决服装的稀疏性问题；
- [0059] 4、该模型可以有效对多个服装间的搭配偏好进行兼容性建模。

### 附图说明

- [0060] 构成本申请的一部分的说明书附图用来提供对本申请的进一步理解,本申请的示意性实施例及其说明用于解释本申请,并不构成对本申请的不当限定。
- [0061] 图1是本发明的流程图；
- [0062] 图2是本发明中用到的商品多模态信息,包含服装的图片,类别层次结构和标题描述；
- [0063] 图3(a)是最流行的上衣和下衣类别搭配；
- [0064] 图3(b)是最流行的下衣和鞋子类别搭配。

### 具体实施方式

- [0065] 应该指出,以下详细说明都是例示性的,旨在对本申请提供进一步的说明。除非另有指明,本文使用的所有技术和科学术语具有与本申请所属技术领域的普通技术人员通常理解相同含义。
- [0066] 需要注意的是,这里所使用的术语仅是为了描述具体实施方式,而非意图限制根据本申请的示例性实施方式。如在这里所使用的,除非上下文另外明确指出,否则单数形式也意图包括复数形式,此外,还应当理解的是,当在本说明书中使用术语“包含”和/或“包括”时,其指明存在特征、步骤、操作、器件、组件和/或它们的组合。
- [0067] 主要包括以下内容：
- [0068] 通过深度卷积神经网络和词袋模型挖掘商品的多模态信息(即视觉模态和文本模态)；
- [0069] 通过多重自编码器神经网络学习多模态商品潜在的兼容性空间；
- [0070] 利用兼容性商品之间的隐式反馈,进一步采用贝叶斯个性化排名框架来挖掘互补商品之间的搭配偏好。
- [0071] 提出基于内容的多模态贝叶斯个性化互补服装搭配框架有效解决稀疏性的问题,并联合地对商品的不同模态关系和商品间隐含的偏好进行兼容性建模。
- [0072] 作为本发明的第一个实施例,提供了基于神经网络的多模态互补服装搭配方法；
- [0073] 基于神经网络的多模态互补服装搭配方法,包括：
- [0074] 步骤(1):从服装的图片中获取视觉特征,同时,从服装的文字描述中获取文本特征；
- [0075] 作为本发明的进一步改进,步骤(1)中：
- [0076] 所述服装,包括:上衣、下衣和鞋子；
- [0077] 所述服装的图片,是指上衣、下衣或鞋子的彩色图片；
- [0078] 所述服装的文字描述,包括:服装的样式、功能和类别；
- [0079] 所述视觉特征,包括:上衣的视觉特征、下衣的视觉特征和鞋子的视觉特征；
- [0080] 所述文本特征,包括:上衣的文本特征、下衣的文本特征和鞋子的文本特征；



[0081] 所述从服装的图片中获取视觉特征,是通过深度卷积神经网络从服装的图片中获取视觉特征;

[0082] 所述从服装的文字描述中获取文本特征,是通过词袋模型从服装的文字描述中获取文本特征;

[0083] 步骤(2):利用自编码器学习不同服装的视觉特征和文本特征的兼容性空间,得到视觉特征的隐含表示和文本特征的隐含表示;

[0084] 作为本发明的进一步改进,步骤(2)中:

[0085] 将上衣 $t_i$ 的视觉特征 $v_i^t$ 、上衣 $t_i$ 的文本特征 $c_i^t$ 、下衣 $b_j$ 的视觉特征 $v_j^b$ 、下衣 $b_j$ 的文本特征 $c_j^b$ 、鞋子 $s_k$ 的视觉特征 $v_k^s$ 和鞋子 $s_k$ 的文本特征 $c_k^s$ 输入到自编码器的编码器中;输出上衣 $t_i$ 的视觉隐含表示 $\hat{v}_i^t$ 、上衣 $t_i$ 的文本隐含表示 $\hat{c}_i^t$ 、下衣 $b_j$ 的视觉隐含表示 $\hat{v}_j^b$ 、下衣的文本隐含表示 $\hat{c}_j^b$ 、鞋子 $s_k$ 的视觉隐含表示 $\hat{v}_k^s$ 和鞋子 $s_k$ 的文本隐含表示 $\hat{c}_k^s$ ;

[0086] 步骤(3):利用解码器将步骤(2)得到的视觉特征的隐含表示和文本特征的隐含表示解码为重建向量;建立重建向量分别与输入特征之间的关系模型;

[0087] 作为本发明的进一步改进,步骤(3)中:

[0088] 通过自编码器的解码器,将上衣 $t_i$ 的视觉隐含表示 $\hat{v}_i^t$ 、上衣 $t_i$ 的文本隐含表示 $\hat{c}_i^t$ 、下衣 $b_j$ 的视觉隐含表示 $\hat{v}_j^b$ 、下衣的文本隐含表示 $\hat{c}_j^b$ 、鞋子 $s_k$ 的视觉隐含表示 $\hat{v}_k^s$ 、鞋子 $s_k$ 的文本隐含表示 $\hat{c}_k^s$ 解码为上衣 $t_i$ 的视觉重建向量 $\hat{v}_i^t$ 、上衣 $t_i$ 的文本重建向量 $\hat{c}_i^t$ 、下衣 $b_j$ 的重建向量 $\hat{v}_j^b$ 、下衣的重建向量 $\hat{c}_j^b$ 、鞋子 $s_k$ 的视觉重建向量 $\hat{v}_k^s$ 、鞋子 $s_k$ 的文本重建向量 $\hat{c}_k^s$ ;

[0089] 作为本发明的进一步改进,步骤(3)中:建立重建向量与输入特征之间的关系模型的步骤为:

$$l_{AE}(x) = l(v_x) + l(c_x)$$

$$= \frac{1}{2} \|\hat{v}_x - v_x\|_2^2 + \frac{1}{2} \|\hat{c}_x - c_x\|_2^2;$$

[0091] 其中, $l_{AE}(x)$ 表示服装 $x$ 重建向量与视觉特征和文本特征之间的关系模型; $l(v_x)$ 表示服装 $x$ 的视觉重建向量与视觉特征的重建误差; $l(c_x)$ 表示服装 $x$ 的文本重建向量与文本特征的重建误差; $\hat{v}_x$ 表示服装 $x$ 的视觉重建向量; $v_x$ 表示服装 $x$ 的视觉特征; $\hat{c}_x$ 表示服装 $x$ 的文本重建向量; $c_x$ 表示服装 $x$ 的文本特征;

[0092] 建立整个服装的重建向量与输入特征之间的关系模型:

$$l_{AE} = l_{AE}(t_i) + l_{AE}(b_j) + l_{AE}(s_k)。$$

[0094] 步骤(4):基于步骤(2)得到的视觉特征的隐含表示和文本特征的隐含表示,建立服装兼容性模型;然后,基于服装兼容性模型,采用贝叶斯个性化排序算法构建兼容性偏好模型;

[0095] 作为本发明的进一步改进,步骤(4)中:建立服装兼容性模型 $comp_{ijk}$ :

$$comp_{ijk} = comp_{ik} + comp_{jk};$$

$$comp_{ik} = (1 - \beta)(\hat{v}_i^t)^T \hat{v}_k^s + \beta(\hat{v}_i^t)^T \hat{v}_k^s;$$

$$comp_{jk} = (1 - \beta)(\hat{v}_j^b)^T \hat{v}_k^s + \beta(\hat{v}_j^b)^T \hat{v}_k^s。$$

[0099] 其中,  $\text{comp}_{ik}$  表示上衣和鞋子的兼容性模型;  $\text{comp}_{jk}$  表示下衣和鞋子的兼容性模型;

[0100] 作为本发明的进一步改进, 步骤 (4) 中: 基于服装兼容性模型, 采用贝叶斯个性化排序算法构建兼容性偏好模型  $l_{bpr}$ :

$$[0101] \quad l_{bpr} = \sum_{(i,j,k_+,k_-)} -\ln \left( s(\text{comp}_{ijk_+} - \text{comp}_{ijk_-}) \right);$$

[0102] 其中,  $\text{comp}_{ijk_+}$  表示上衣和下衣对正例鞋子的兼容性偏好,  $\text{comp}_{ijk_-}$  表示上衣和下衣对负例鞋子的兼容性偏好;  $s(\text{comp}_{ijk_+} - \text{comp}_{ijk_-})$  表示神经网络的阈值函数;  $(i, j, k_+, k_-)$  表示上衣  $t_i$  和下衣  $b_j$  相对于鞋子  $s_{k_-}$  来说更适合与鞋子  $s_{k_+}$  搭配。

[0103] 步骤 (5): 基于步骤 (2) 得到的视觉特征的隐含表示和文本特征的隐含表示, 建立视觉特征隐含表示和文本特征隐含表示的一致性模型; 然后, 建立服装的多模态隐含特征一致性模型;

[0104] 作为本发明的进一步改进, 步骤 (5) 中: 视觉特征隐含表示和文本特征隐含表示的一致性模型  $l_{vc}(t_i)$ :

$$[0105] \quad l_{vc}(t_i) = -\ln \left( s((\tilde{v}_i^t)^T \tilde{c}_i^t) \right);$$

[0106] 其中,  $s(\tilde{v}_i^t \tilde{c}_i^t)$  表示神经网络的阈值函数;  $\tilde{v}_i^t$  表示服装  $t_i$  的视觉特征隐含表示;  $\tilde{c}_i^t$  表示服装  $t_i$  的文本特征隐含表示。

[0107] 作为本发明的进一步改进, 步骤 (5) 中: 服装的多模态隐含特征一致性模型  $l_{mod}$ :

$$[0108] \quad l_{mod} = \sum_{(i,j,k_+,k_-)} l_{vc}(t_i) + l_{vc}(b_j) + l_{vc}(s_{k_+}) + l_{vc}(s_{k_-});$$

[0109] 其中,  $l_{vc}(t_i)$  表示上衣视觉特征隐含表示和文本特征隐含表示的一致性模型;  $l_{vc}(b_j)$  表示下衣视觉特征隐含表示和文本特征隐含表示的一致性模型;  $l_{vc}(s_{k_+})$  表示正例鞋子视觉特征隐含表示和文本特征隐含表示的一致性模型;  $l_{vc}(s_{k_-})$  表示负例鞋子视觉特征隐含表示和文本特征隐含表示的一致性模型。

[0110] 步骤 (6): 基于步骤 (3)、步骤 (4) 和步骤 (5) 的计算结果, 构建基于深度神经网络的多模态互补服装搭配模型; 对已经构建的多模态互补服装搭配模型进行训练; 利用已经训练好的多模态互补服装搭配模型进行服装搭配推荐。

[0111] 作为本发明的进一步改进, 步骤 (6) 中: 基于深度神经网络的多模态互补服装搭配模型:

$$[0112] \quad L = l_{AE} + l_{mod} + l_{bpr};$$

[0113] 作为本发明的进一步改进, 步骤 (6) 中: 对已经构建的多模态互补服装搭配模型参数通过随机梯度下降方法进行训练, 通过迭代至该模型收敛, 输出最终参数。

[0114] 作为本发明的进一步改进, 步骤 (6) 中: 利用已经训练好的基于深度神经网络的多模态互补服装搭配进行服装搭配推荐:

[0115] 通过训练好的最终参数, 计算所有  $\text{comp}_{ijk}$  的值;

[0116] 其中  $\text{comp}_{ijk}$  最大的值所对应的鞋子即为上衣和下衣最为搭配的鞋子。

[0117] 作为本发明的第二个实施例, 提供了基于神经网络的多模态互补服装搭配系统;

[0118] 基于神经网络的多模态互补服装搭配系统, 包括: 存储器、处理器以及存储在存储

器上并在处理器上运行的计算机指令,所述计算机指令被处理器运行时,完成上述任一方法所述的步骤。

[0119] 作为本发明的第三个实施例,提供了一种计算机可读存储介质;

[0120] 一种计算机可读存储介质,其上存储有计算机指令,所述计算机指令被处理器运行时,完成上述任一方法所述的步骤。

[0121] 本发明的方案如图1所示,下面结合具体的例子对本发明中的技术方案进行描述:

[0122] S1:首先我们通过图2展示了几个服装的样例,每个样例都包含服装的图片,服装的所属类别和标题描述。其中服装的颜色、图案等信息可以从图片中直观的展示,而服装的样式、功能和类别等信息可以从文本信息中有效的获得。且不同的模态,即视觉模态和文本模态,可以从不同的方面来表示同一个物品。

[0123] S2:我们进一步探索了文本信息的有效性,图3(a)为最流行的上衣和下衣的类别搭配,其中每一个圆点代表一个类别,且浅灰色代表上衣类别,深灰色表示下衣类别,圆的面积和线的宽度分别与相应类别的服装数量和不同类别相搭配的同现频率成正比。我们可以看到毛衣、T恤和及膝长裙是与其它类别搭配较多的服装,且我们还可以看到外套和连衣裙更搭,而毛衣则和及膝长裙搭配较多。同样,从图3(b)最流行的下衣和鞋子搭配类别中我们可以看出,及膝长裙和高跟鞋比较搭,而脚踝靴和紧身牛仔裤较搭。

[0124] S3:通过由caffe提供的预训练的Alexnet网络进行视觉特征提取,它包含五个卷积层和3个全连接层。我们提取Alexnet网络的‘fc7’层的输出结果作为视觉特征,最终我们得到一个4096维的特征,对于上衣 $t_i$ 、下衣 $b_j$ 和鞋子 $s_k$ ,其视觉特征可分别表示为 $v_i^t, v_j^b, v_k^s$ 。

[0125] S4:通过物品类别和标题描述构建词汇表并进行过滤。然后通过词袋模型,我们分别得到上衣 $t_i$ 、下衣 $b_j$ 和鞋子 $s_k$ 的文本特征 $c_i^t, c_j^b, c_k^s$ ,其特征均为3345维。

[0126] S5:本发明通过学习一个潜在的兼容性空间来弥合异构空间中物品的语义鸿沟,而不是直接通过上述的特征向量来直接计算异构空间中上衣、下衣和鞋子的兼容性。这个潜在的兼容性空间可以让不同种类的物品具有最大的兼容性,例如上衣、下衣和鞋子具有不同的样式和功能等。该发明可以学习一个兼容性空间,使得不同种类的服装从颜色、图案、风格以及材料等因素上找到其一致性。这里我们通过一个多重的自编码器来学习上衣 $t_i$ 、下衣 $b_j$ 和鞋子 $s_k$ 的视觉特征 $v_i^t, v_j^b, v_k^s$ 和文本特征 $c_i^t, c_j^b, c_k^s$ 的兼容性空间。通过自编码器的编码器,我们分别得到上衣、下衣和鞋子的视觉和文本隐含表示 $\tilde{v}_i^t, \tilde{c}_i^t, \tilde{v}_j^b, \tilde{c}_j^b, \tilde{v}_k^s, \tilde{c}_k^s$ 。通过自编码器的解码器,我们将这些隐含表示解码为重建表示 $\hat{v}_i^t, \hat{c}_i^t, \hat{v}_j^b, \hat{c}_j^b, \hat{v}_k^s, \hat{c}_k^s$ 。对于物品 $x$ ,重建后的向量应与输入时的向量相接近,因此我们最小化下面的模型来进行学习:

$$l_{AE}(x) = l(v_x) + l(c_x)$$

$$[0127] \quad = \frac{1}{2} \|\hat{v}_x - v_x\|_2^2 + \frac{1}{2} \|\hat{c}_x - c_x\|_2^2$$

[0128] 本发明中我们同时对上衣 $t_i$ 、下衣 $b_j$ 和鞋子 $s_k$ 进行学习:

$$[0129] \quad l_{AE} = l_{AE}(t_i) + l_{AE}(b_j) + l_{AE}(s_k)$$

[0130] S6:对物品的兼容性进行建模,这里以为上衣和下衣搭配鞋子为例:通过学习一个潜在的兼容性空间,我们分别得到上衣、下衣和文本的视觉和文本隐含表示  $\tilde{v}_i^t, \tilde{c}_i^t, \tilde{v}_j^b, \tilde{c}_j^b, \tilde{v}_k^s, \tilde{c}_k^s$ 。基于这些隐含表示,我们得到上衣、下衣和鞋子的兼容性模型:

$$[0131] \quad \text{comp}_{ijk} = \text{comp}_{ik} + \text{comp}_{jk}$$

[0132] 其中,  $\text{comp}_{ik} = (1 - \beta)(\tilde{v}_i^t)^T \tilde{v}_k^s + \beta(\tilde{v}_i^t)^T \tilde{v}_k^s$ ,同理可得到下衣和鞋子的兼容性  $\text{comp}_{jk}$ 。

[0133] S7:为充分挖掘物品的视觉和文本模态,我们对商品  $x_i$  的视觉和文本隐含特征的一致性进行建模:

$$[0134] \quad l_{vc}(t_i) = -\ln\left(s((\tilde{v}_i^t)^T \tilde{c}_i^t)\right)$$

[0135] 同时考虑上衣  $t_i$ 、下衣  $b_j$  和鞋子  $s_k$ ,我们可以得到三者的多模态隐含特征一致性模型:

$$[0136] \quad l_{mod} = \sum_{(i,j,k_+,k_-)} l_{vc}(t_i) + l_{vc}(b_j) + l_{vc}(s_{k_+}) + l_{vc}(s_{k_-})$$

[0137] S8:为充分挖掘上衣、下衣和鞋子在搭配中的兼容性,我们采用贝叶斯个性化框架,以为上衣和下衣推荐鞋子为例,其模型如下:

$$[0138] \quad l_{bpr} = \sum_{(i,j,k_+,k_-)} -\ln\left(s(\text{comp}_{ijk_+} - \text{comp}_{ijk_-})\right)$$

[0139] 其中  $\text{comp}_{ijk_+}$  和  $\text{comp}_{ijk_-}$  表示上衣和下衣对不同鞋子的兼容性偏好,即上衣  $t_i$  和下衣  $b_j$  相对于鞋子  $s_{k_-}$  来说更适合与鞋子  $s_{k_+}$  搭配。

[0140] S9:最终我们得到基于深度神经网络的多模态互补服装搭配模型:

$$[0141] \quad L = l_{AE} + l_{mod} + l_{bpr}$$

[0142] 该发明通过训练该模型,最终得到上衣、下衣和鞋子的多模态兼容性空间,其中  $\text{comp}_{ijk}$  最大的值所对应的鞋子即为上衣和下衣最为搭配的鞋子。

[0143] 以上所述仅为本申请的优选实施例而已,并不用于限制本申请,对于本领域的技术人员来说,本申请可以有各种更改和变化。凡在本申请的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本申请的保护范围之内。

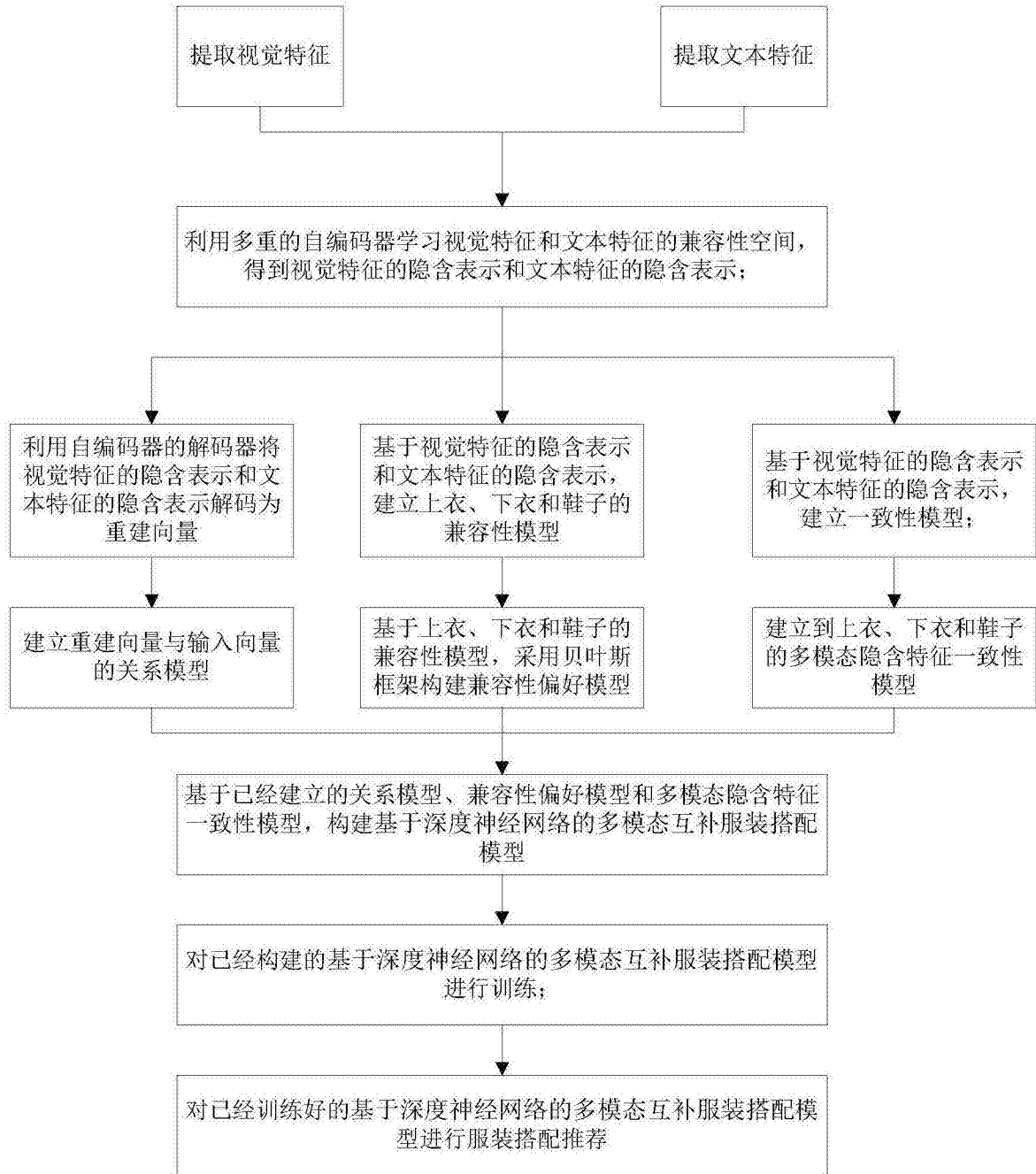


图1

样例	图片	类别	标题
1		Women's Fashion > Clothing > Tops > Shirts	Red Plaid Long Sleeved Shirt
2		Women's Fashion > Skirts > Knee Length Skirts	Floral pleated full skirt
3		Women's Fashion > Shoes > Pumps	Brown Pointed Toe Single Strap Over Ladies Heels

图2

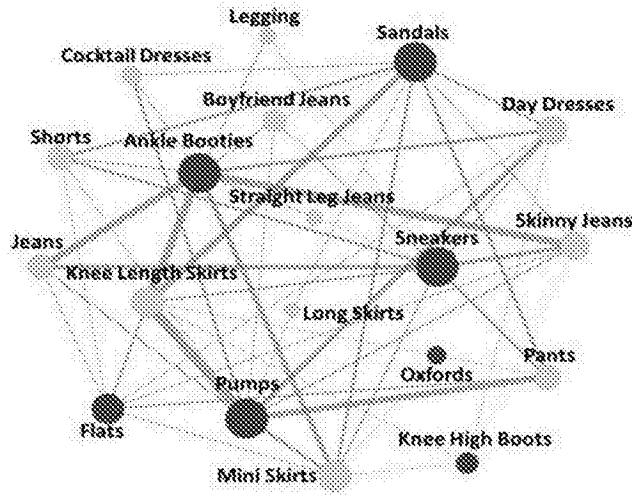


图3 (a)

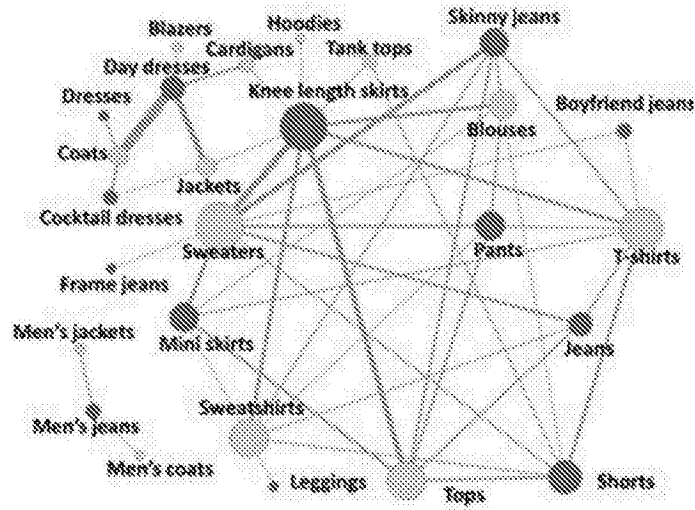


图3 (b)