(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 103345645 B
(45)授权公告日 2016.09.28

(21)申请号 201310262308.9

(22)申请日 2013.06.27

(65)同一申请的已公布的文献号
申请公布号 CN 103345645 A

(43)申请公布日 2013.10.09

(73)专利权人 复旦大学
地址 200433 上海市杨浦区邯郸路220号

(72)发明人 张玥杰 张溢 金城 薛向阳

(74)专利代理机构 上海正旦专利代理有限公司 31200
代理人 陆飞 盛志范

(51)Int.Cl.
G06K 9/66(2006.01)
G06K 9/46(2006.01)

(56)对比文件
CN 102663446 A,2012.09.12,
CN 102663010 A,2012.09.12,
CN 102254043 A,2011.11.23,
CN 101561814 A,2009.10.21,
CN 101866339 A,2010.10.20,
CN 102364497 B,2013.06.05,
JP 2002183519 A,2002.06.28,
CN 101794309 A,2010.08.04,
赵春晖等.一种基于词袋模型的图像优化分类方法.《电子与信息学报》.2012,第34卷(第9期),
尹达等.基于颜色特征的图像检索技术研究.《计算机系统应用》.2009,(第5期),第151-155页.
路斌等.一种基于SVM的多层分类策略.《软件技术与数据库》.2005,第31卷(第1期),第73-75页和第113页.
李飒.基于内容的图像检索技术在多类别商品图像检索中的应用.《万方数据库硕士学位论文》.2010,全文.

审查员 袁冠群

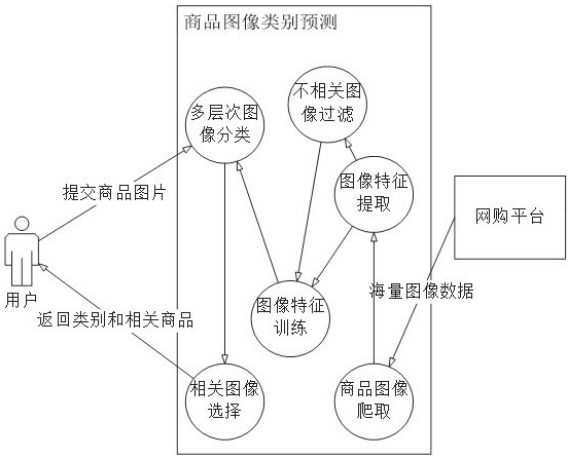
权利要求书5页 说明书17页 附图4页

(54)发明名称

面向网购平台的商品图像类别预测方法

(57)摘要

本发明属于多媒体信息检索技术领域,具体为基于网购平台的商品图像类别预测方法。本发明主要包含六个模块及相关算法,即训练图像的获取,图像特征提取,不相关图像过滤,图像特征训练,多层次图像分类,相关图像选择。本发明基于从网购平台上获取的真实数据,通过大规模数据的训练,能够自动分析图像中商品的类别信息,向用户提供购物指引,从而简化用户在线购物流程,增强用户体验,在图像检索领域具有广泛的应用价值。



CN 103345645 B

1.一种基于网购平台的商品图像类别预测方法，其特征在于具体步骤如下：

（1）获取训练图像，向当前的网购平台爬取商品图像和图像相关标注，并初步清理垃圾数据，为训练图像分类模型提供数据；

（2）提取图像特征，选择特定的特征表达方法，将爬取的图像从点阵表示转化为特征表示；

（3）过滤不相关图像，利用步骤(2)中所产生的特征表达，将与标注不相关的商品图像去除；

（4）训练图像特征，对于图像的兴趣点特征表达，进一步训练BOW词典，将图像转化为词包表达；

（5）多层次图像分类，利用图像的BOW特征，训练多层次的图像分类模型，并应用于用户输入图像的类别预测；

（6）选择相关图像，根据步骤(5)中所提供的类别预测，选取返回给用户的相关图像；

其中，所述提取图像特征，首先，选取不同的图像特征，并对特征间的相似性进行定义；其中，图像特征包含颜色、纹理和兴趣点特征三部分；

抽取颜色特征，首先将原始图像按照三种不同的尺度划分为共 $\sum_{r=0}^{2} 2^r \times 2^r = 21$ 个网格，并对每个网格抽取基于36个色彩饱和度明暗空间的颜色直方图特征，共 $21 \times 36 = 756$ 维颜色直方图特征；基于这些颜色直方图特征，两幅图像u和v之间颜色直方图的相似性 $k_c(u,v)$ 定义为：

$$\kappa_c(u,v) = \frac{1}{\sum_{r=0}^{R-1} \frac{1}{2^{R-r+1}}} \sum_{r=0}^{R-1} \frac{1}{2^{R-r+1}} N_r(u,v) \quad (1)$$

其中，R=3，是所取网格划分图像尺度的种类数量；$N_0(u,v)$ 表示原始分辨率的颜色直方图相似性；$N_r(u,v)$ 表示第r种分辨率的颜色直方图相似度；

$$N_r(u,v) = \sum_{i=1}^{36} Norm\left(H_i^r(u), H_i^r(v)\right) \quad (2)$$

其中，$H_i^r(u)$ 和 $H_i^r(v)$ 分别表示图像u和v中，第r种分辨率网格划分中第i 个格子的颜色直方图相似度；Norm代表的是二阶标准距离；

抽取纹理特征，三个尺度四个方向共12个21×21像素点的Gabor滤波器被分别使用于对图像做滤波操作；分别计算滤波后12幅图像所有像素点的均值和方差，得到 $12 \times 2 = 24$ 维的Gabor纹理特征；

基于上述纹理特征，两幅图像u和v之间Gabor纹理的相似性 $\kappa_t(u,v)$ 定义为：

$$\kappa_t(u,v) ::: e^{-d_t(u,v)/\sigma_t}$$

$$d_t(u,v) = Norm\left(g_i(u), g_j(v)\right) \quad (3)$$

其中，σ_t 代表所有图像 $d_t(u,v)$ 的均值；$g_i(u)$ 和 $g_j(v)$ 分别代表图像u的第i个和图像v的

第j个Gabor滤波器；

抽取兴趣点特征，选用SURF算法对图像进行处理；SURF算法提取的每个兴趣点有64维；

将两幅图像的兴趣点做一一配对，使得所有兴趣点配对间二阶标准距离之和最小；该配对用二分图匹配算法实现；于是，两幅图像u和v之间SURF特征的相似性$\kappa_s(u,v)$定义为：

$$\kappa_s(u,v) = e^{-d_s(u,v)/\sigma_s}$$

$$d_s(u,v) = \frac{\sum_i Norm(s_i(u), s_i(v))}{\sum i} \qquad (4)$$

其中，σ_s代表所有图像$d_s(u,v)$的均值；$s_i(u)$和$s_i(v)$分别代表图像u的第i个兴趣点和图像v中与其配对的兴趣点；

最后，视觉相似性通过一个混合的线性加权统计出最终的结果，因此，图像u和v之间的视觉相似性定义为：

$$\kappa(u,v) = \beta_1\kappa_c(u,v) + \beta_2\kappa_s(u,v) + \beta_3\kappa_t(u,v)$$

$$\sum_{i=1}^{3}\beta_i = 1 \qquad (5)$$

其中，β_i表示每一种特征所占的权重，根据所有图像间$\kappa_c(u,v)$、$\kappa_s(u,v)$和$\kappa_t(u,v)$的方差进行分配；将所有的特征的相似性合并起来，将图像间的相似性问题简化，使后续应用更易处理。

2.根据权利要求1所述的预测方法，其特征在于所述过滤不相关图像，是利用图像预先定义的特征和相似性，对不相关图像进行过滤；

首先对图像聚类，定义图像类与图像类之间的类间距离以及单个图像类内部的类内距离：

对于两个图像类G_i和G_j而言，它们的类间距离定义为：

$$s(G_i, G_j) = \frac{\sum_{u\in G_i}\sum_{v\in G_j}\kappa(u,v)}{\sum_{u\in G_i}1 \times \sum_{v\in G_j}1} \qquad (6)$$

对于图像类G_i而言，其类内距离相应的定义为：

$$s(G_i, G_i) = \frac{\sum_{u\in G_i}\sum_{v\in i(u\neq v)}\kappa(u,v)}{\sum_{u\in G_i}1 \times \sum_{v\in G_i}1} \qquad (7)$$

对于类内距离大于所有与其他类的类间距离的图像类，将其再度划分；对于两类类间距离小于两类类内距离的，则将两类合并，通过这样两条规则，结合标准割算法，对图像分类进行不断迭代；当迭代次数达到一定值或类别数量达到预设最大类别数量K时，停止迭代；这时，从结果中选取其中图像数量最多的类，将类中的所有图像作为相关图像，而将其他类别的图像作为不相关图像。

3.根据权利要求2所述的预测方法，其特征在于所述的对于图像的兴趣点特征表达，进一步训练BOW词典，就是利用层次聚类和K-means本身结合的方法用于K-means算法的初始

点选择,这种被称为层次K-means聚类的算法具体流程如下:

(1)设置迭代次数$i = 0$;

(2)利用K-Means算法对原始数据进行聚类,K个随机点作为原始聚类中心,算法达到收敛条件后得到K个聚类中心$\{C_k^i\}$;

(3)当i<M,M为预设最大迭代次数,i=i+1,重复执行流程(2);

(4)将得到的所有M * K个聚类中心作为样本点($g_j = C_k^i, j = i * K + k$),执行基于重心距离的层次聚类算法:

(a)该算法将所有初始样本看作类中心$C_j = g_j$(1<=j<=N * K)

(b)计算任意两类中心之间的距离作为类与类的距离,将距离最小的两类合并:

$$\delta(C_S, C_T) = \mathrm{d}(\mathrm{cs}, \mathrm{ct})$$

$$\mathrm{vs} = \frac{1}{|C_S|}\sum g_j \in C_S, \quad vt = \frac{1}{|C_T|}\sum g_j \in C_T \qquad (8)$$

(c)重复执行(b)直至最终只剩下K类;

(5)以流程(4)中得到的K个类中心作为初始类中心,执行按照流程(1)-(3)K-means算法,直至收敛,得到最终的K个类中心。

4.根据权利要求3所述的预测方法,其特征在于对于基于图像的兴趣点表达,训练视觉BOW词典,进行进一步优化,具体是通过利用每个样本点与其上一轮所分配中心的距离和三角形不等式模型,推测其与本轮所有中心的距离关系;首先定义相关变量如下:

x_i	第 i 个样本点的向量表示
a_i	第 i 个样本点当前所属中心的编号
$disx_i$	第 i 个样本到其当前所属中心的距离
c_j	当前一轮待分配的第 j 个中心的向量表示
c_j'	当前一轮待分配的第 j 个中心的向量表示
$mindisc_j$	当前一轮待分配的第 j 个中心与其最近中心的距离
$mindisc_j'$	上一轮第 j 个待分配中心和当前待分配中心的最小距离
$secmindisc_j'$	上一轮第 j 个待分配中心和当前待分配中心的第二小距离
$disc_j$	第 j 个中心向量这一轮和上一轮的距离
u_i	当前一轮第 i 个样本点到其最近中心的上界
l(i,j)	当前一轮第 i 个样本点到第 j 个待分配中心距离的下界
d(a,b)	a, b 连个向量的距离

根据上述定义，利用3个三角形不等式优化相关的距离计算，其迭代过程中判断样本点所属中心的关键执行步骤如下：

(1)若$2*disx_i+mindisc_j'<secmindisc_j'$成立，则第i个样本i直接分配给中心$a_i$，否则执行步骤(2)；

(2)若$2*(disc_{a_i}+disx_i)<mindisc_i$成立，则第i个样本点直接分配给中心$a_i$，否则$u_i=d(x_i,c_{a_i})$；

(3)若$2*u_i=d(c_{a_i},c_j)$成立，则第i个样本点至中心a_i的距离小于其与第j个中心点的距离，可省去其与第j个中心点的距离计算；在步骤(2)，(3)均不满足的条件下，需要计算第i个样本点与第j个中心点的距离，更新u_i。

5.根据权利要求3或4所述的预测方法，其特征在于所述的利用图像的BOW特征，训练多层次的图像分类模型，是将基于SVM分类方法的算法用于训练分类模型；为解决BOW特征的稀疏性问题选取一种改进的RBF核——X^2-RBF核作为SVM核函数，该核函数的定义为：

$$K(x,z)=exp\left(-\frac{(x-z)^2}{2\sigma^2(x+z)}\right) \qquad (9)$$

利用商品类别本身的层次属性，从商品类别的最高层开始，自上而下地对商品的类别进行预测，这种层次分类的方法将商品的分类关系表达成树结构，当树的节点具有多个子节点时训练一个多类分类问题的模型，这种树结构的关系定义为：

$$\forall c_i,c_j\in C,\text{如果 } c_i\prec c_j \text{ 那么} c_j\not\prec c_i$$

$$\forall c_i\in C,c_i\not\prec c_i \qquad (10)$$

$$\forall c_i,c_j,c_l\in C,\text{如果 } c_i\prec c_j \text{ 并且} c_j\prec c_l \text{可以推出} c_i\prec c_l$$

其中，c_i,c_j,c_l分别代表第i，j，l个类别，C表示所有类别的集合；

在这样的树结构中，存在多种类别划分的策略用于层次分类，相关的变量定义如下：

Tr	所有训练样本
$Tr^+(c_j)$	对于类别 c_j 而言的所有正样本
$Tr^-(c_j)$	对于类别 c_j 而言的所有负样本
$\uparrow(c_j)$	c_j 的父类别
$\downarrow(c_j)$	c_j 的所有子类别集合
$\Uparrow(c_j)$	c_j 的所有祖先类别集合
$\Downarrow(c_j)$	c_j 的所有子孙类别集合
$\leftrightarrow(c_j)$	c_j 的所有兄弟类别集合
$*(c_j)$	c_j 中所有样本点的集合

基于相应的定义，选用如下方法定义正负样本：

$$Tr^{+}(c_j) = *(c_j) \cup \Downarrow (c_j), Tr^{-}(c_j) = Tr \backslash *(c_j) \cup \Downarrow (c_j) \cup \Uparrow (c_j) \quad (11)$$

这种定义方式通过自顶向下的顺序，对叶子节点类别进行分类模型训练；每次分类模型的训练只包含同一父亲节点的所有兄弟节点；选取一对一的算法，解决该小规模的多类分类问题，经过自顶向下，3-4次小规模多类别的分类之后，即得到样本的最终类别预测。

6.根据权利要求5所述的预测方法，其特征在于所述的利用图像的BOW特征，训练多层次的图像分类模型，在所述层次分类的基础上，加入一些潜在可能分类，使高层误分类情况能够得到缓解，其具体步骤如下：

(1)在最高层的类别中，根据一对一算法预测时的排序结果，选择前五个类别作为商品图像备选的类别；

(2)分别将商品图像应用于上一步所产生的五个类别中，亦根据一对一算法每类产生五个子类别，得到25个相对于上一步中孙子代的备选类别；

(3)为步骤(2)中的25个类别训练一对一的多类SVM分类模型，根据其投票机制，选取排名前五的类别循环执行步骤(2)，直至所得到的五个类别均为叶子类别。

7.根据权利要求1所述的预测方法，其特征在于步骤(7)所述的从网购平台爬取图像用于选取分类模型训练数据的过程中做如下处理：

(1)在将爬取的商品图像用于分类训练前，在爬取原始商品图像时，按照预计训练图像的两倍以上的规模爬取；

(2)在从网购平台爬取商品图像时，按照平台所提供统一的规格图像进行爬取；

(3)在应用SURF算法提取特征时，尺寸过小的图像和长宽比例极不协调的图像将会无法提取，对于商家提供的尺寸过小的图像和长宽比例极不协调的图像在爬取过程中避免；

(4)所有类别需要保证在类别树中的深度一致。

8.一种基于权利要求7所述预测方法的系统，其特征在于包括如下6个模块：训练图像的获取模块，图像特征提取模块，不相关图像过滤模块，图像特征训练模块，多层次图像分类模块，相关图像选择模块。

面向网购平台的商品图像类别预测方法

技术领域

[0001] 本发明属于多媒体信息检索技术领域，具体涉及一种商品图像类别预测方法。

背景技术

[0002] 在互联网在线购物领域，数字图像信息有着文本信息不可取代的地位。尤其是在个人对个人(Consumer to Consumer, C2C)和商家对顾客(Business to Customer, B2C)这类应用当中，消费者存在迫切地希望能够看到商品的真实外观的需求。然而，相比文本信息，数字图像信息在计算机中存储和传输所占用和消耗的资源都要大得多，这导致早期互联网对图像信息的使用非常谨慎。幸运的是，随着计算机技术和互联网技术的高速发展，限制数字图像甚至高质量的数字图像内容在互联网中存储和传输的瓶颈已经得到极大缓解。另一方面，近年来随着物流领域的逐渐成熟和人们观念的转变，在线购物也逐渐成为人们购物的主要渠道之一，网购平台在这样的环境下已经取得了长足的发展。在这种背景下，如淘宝、京东和亚马逊等网购平台已经积累大量的商品图像信息，对于这些平台而言，如何更有效地实现对数字图像信息的组织、分析、检索和向消费者展示已经变得十分重要[1]。

[0003] 在网购平台网站中，商品图像的标题和分类等信息可以看成是商品图像的附属标签信息。合理地利用这些标签能够指引用户根据自己需求浏览内容[2]，可以提升消费者的使用体验，成为消费者浏览网购平台的重要辅助手段。在这种前提下，对商品图像类别的预测，不论是于上传商品图像的商户而言还是对浏览商品图像的用户而言都是有着重大意义的。然而要实现对于商品图像类别的预测，在当前的网购平台上，还存在着诸多的挑战。

[0004] 首先，网购平台上的商品图像附属类别标签信息是由个体商户所提供的。同其他社会化的多媒体数字图像分享平台一样，这些上传者可以认为是社会化的上传者。因此，这些标签信息往往存在着与图像间不相关的情况[3]。这种相关情况取决于多个方面：

[0005] (1)网购平台上不存在相关的类目。随着网购平台的发展壮大，这种情况正在不断减少。并且，大多数网购平台的类目是层次结构的，因此即使没有准确的类目，也会有相关的高层类目或在这些高层类目所包含的其他类别中。另一方面，商品图像的标题信息一般可以自由添加，在这个方面不存在限制。

[0006] (2)在附属标签的添加者和商品检索者之间存在语义鸿沟[4]。所谓语义鸿沟，一般是指不同用户之间对图像的视觉表现理解是不同的。而在精确的商品图像检索过程中，这种鸿沟更进一步体现为不同的用户对于相同商品名称表述的区别和对于不同商品名称表述的混淆。这类问题在中国这个幅员辽阔的国家更为明显。不同地区、不同民族有着不同的方言，在不同方言中，对于商品的名称往往有不同的表述。对于这个问题，许多商品图像的上传者会通过添加商品名称的多个表述作为商品图像的标签，但这种处理方式本身对特定的商品检索用户而言会带来不相关的标签，甚至带有误导作用的标签。

[0007] (3)商品图像排序规则引起的过度优化行为。在网购平台上，商家为了牟利，希望自己的商品能得到更多的曝光次数。其最为重要的手段之一就是针对网购平台搜索引擎进行搜索引擎优化(Search Engine Optimization, SEO)。商家往往会选择用户搜索较多的

热门关键词标签，并选择其中与商品相关度较大的标签添加给商品。但在这种情况下，商家选择添加何种标签全凭自身职业道德的约束，因此在竞争激烈的网购平台中，会存在有些商家为了吸引用户，添加与商品相关度并不高的标签的情况发生。

[0008] 因此，要利用网购平台自身的图像，首先需要对商品图像的标签信息进行清理，找出真正存在巨大相关性的标签。在社会化图像分享平台上，这个问题有着较多的研究[5, 6, 7, 8]。传统解决方案是利用人工重新为训练数据集图像标记一些准确的标签，通过这些准确的标签，以及图像的低层次特征，训练这些标签与图像低层次特征之间的相关性模型，最后用这些模型来实现对于图像标签的清理或预测。这类方法的优点是，得到的结果相对准确，但是但其缺点也十分明显，即需要大量的人工标记，这往往会耗费巨大的人力成本，并且对图像本身的社会化标注而言是一种浪费。为了广泛地利用社会化标注，一些研究则将用户标注、图像和图像特征之间建立相应的图关系。例如可以用这三者建立超图，在图模型之上，可以利用图划分算法实现图像与标签之间相关性的计算[9]。也可以将这三者建立一个或多个二分图，利用协同过滤算法，将图像划分到相应的标签上，从而实现清理不相关标签的效果[10, 11, 12]。也有研究从大规模数据的角度出发，采用部分无监督的方法建立图像视觉的语义网络，并利用该语义网络和多模态的信息，对与标签不相关的图像进行过滤[13, 14]。

[0009] 其次，在大规模数据条件下，图像特征的提取也是重大的挑战之一。不论在标签信息清理还是商品图像分类领域，图像特征提取都是这些领域的基础工作。

[0010] 在标签信息清理的问题中，图像信息往往需要用到图像的多种特征。为了适应大规模数据的处理，颜色特征和一些简单的纹理特征是较好的选择[15]。而为了取得更好的效果，尺度旋转不变的兴趣点特征(Scale-Invariant Feature Transform, SIFT)[16]也是相当有用的特征。但在大规模的数据处理的条件下，效果相似，速度更快，且特征维度更低的加速算法(Speeded Up Robust Features, SURF)[17]则是更为合适的选择。

[0011] 在图像分类领域，基于视觉词包(Bag of Visual Words, BOW)的分类算法是最为主流的算法[18]。在图像检索和分类应用中，由于图像的数量和词典的规模巨大，词典的训练速度将成为应用的瓶颈。因此，K-means的聚类方法成为了训练词典的较好方法。但是尽管经典的K-means算法在聚类算法中是一种速度较快的算法，大规模数据的情况下，其执行效率依然会因为大量重复的计算而显得底下。为解决这一相关问题，有学者提出利用三角形不等式加速K-means的方法[19]，在理论上能够为K-means算法加速百倍以上。然而，这种算法在K-means算法每轮的迭代过程中需要存储及其大量的中间数据，使得其难以全部存放在计算机主存中从而导致其实际加速效果在大规模数据中效果并不佳。在此之后，基于这种方法，又有学者提出一些在运行速度和主存空间使用中折中的优化算法[20, 21, 22]。这些算法在词典较小的情况下的执行效率甚至能够超过[19]中所述的算法。

[0012] 最后，大规模数据条件下的图像分类也是商品图像类别预测任务的重大挑战。在这个任务中，大规模数据体现在两个方面。一方面是网购平台中商品图像的数量巨大，对于每一个类别而言，可以用于训练的图像数据极为丰富，充分利用这大规模的数据，使其发挥最大的效果是难点之一；另一方面是商品类别多，随着网购平台的发展，在线购物几乎可以买到所有线下可以购买到的商品，因此商品种类繁多，类别与类别之间的区分越来越小。

[0013] 在图像类别预测领域，传统的方法主要有使用SVM分类器训练金字塔匹配模

型[23]、基于仿生学的启发式模型[24, 25]和直接使用KNN分类的模型[26, 27, 28]等。近年来，也有利用非线性SVM分类器训练空间金字塔(Spatial Pyramid Matching, SPM)的模型[29]在一些知名的图像分类数据集上取得不错的效果。当然，最知名的还是要数基于BOW的分类算法。这些分类算法在小规模的数据集中能够取得较好的效果。但对于当今的商品图像类别预测，由于类别数量极多，所以运算速度非常缓慢，难以应用中直接使用。

[0014] 在类别数量特别多的情况下，基于不同的分类模型，有研究人员利用层次分类的方法对分类应用进行优化。通过对层次的不同定义，层次分类可以应用于不同的分类场合，从而提高分类的准确率和效率[30]。其中，与SVM分类器结合较好的有层次SVM分类[31]和基于贝叶斯方法的SVM分类[32]。这些方法和SVM分类器一样，可以独立于特征，解决普遍的多类别数量的分类问题。

[0015] 由上述分析可以看到，要实现基于网购平台上商品图像类别的预测，主要需要解决的是在大规模图像数据背景下，图像与社会化标注间相关性的衡量、图像特征的提取以及多类别图像层次分类的问题。因此，本发明由图像特征提取、不相关图像过滤、视觉词典训练和多类别图像层次分类四个模块构成。这些模块中的核心算法构成本发明的核心内容。

[0016] 参考文献

[0017] [1]Datta, R., Joshi, D., Li, J., and Wang, J.Z. 2008. Image retrieval: Ideas, influences, and trends of the new age. ACM Computing Surveys (CSUR), 40(2): Article 5.

[0018] [2]Liu, D., Hua, X.S., Yang, L.J., Wang, M., and Zhang, H.J. 2009. Tag ranking. In Proc. of WWW 2009, 351-360.

[0019] [3]Kennedy, L.S., Chang, S.F., and Kozintsev, I.V. 2006. To search or to label: predicting the performance of search-based automatic image classifiers. In Proc. of MIR 2006, 249-258.

[0020] [4]Zhou, N., Peng, J.Y., Feng, X.Y., and Fan, J.P. 2011. Towards more precise social image-tag alignment. In Proc. of MMM 2011, Vol. Part II, 46-56.

[0021] [5]J. Li and J. Z. Wang. 2008. Real-Time Computerized Annotation of Pictures. In IEEE Transactions on Pattern Analysis and Machine Intelligence.

[0022] [6]F. Monay and D. G. Perez. 2003. On Image Auto-annotation with Latent Space Modeling. In Proceeding of 10th ACM International Conference on Multimedia.

[0023] [7]G. Sychay, E. Y. Chang and K. Goh. 2002. Effective Image Annotation via Active Learning. In IEEE International Conference on Multimedia and Expo.

[0024] [8]R. Shi, C. H. Lee and T. S. Chua. 2007. Enhancing Image Annotation by Integrating Concept Ontology and Text-based Bayesian Learning Model. In Proceeding of 14th ACM International Conference on Multimedia.

[0025] [9]Gao, Y., Wang, M., Luan, H.B., Shen, J.L., Yan, S.C., and

Shuicheng Yan, andTao, D.C. 2011. Tag-based social image search with visual-text joint hypergraph learning. In Proc. of ACM MM 2011, 1517-1520.

[0026] [10] G. Qiu. 2004. “Image and Feature Co-clustering”. ICPR (4):991-994.

[0027] [11] B. Gao, T.-Y. Liu, T. Qin, X. Zhang, Q.-S. Cheng, W.-Y. Ma. 2005. “Web image clustering by consistent utilization of visual features and surrounding texts”, ACM Multimedia.

[0028] [12] M. Rege, M. Dong, J. Hua. 2008. “Graph theoretical framework for simultaneously integrating visual and textual features for efficient web image clustering”, WWW.

[0029] [13] Yang, C.L., Peng, J.Y., Feng, XY., and Fan, J.P. 2012. Integrating bilingual search results for automatic junk image filtering. Multimedia Tools and Applications.

[0030] [14] Gao, Y.L., Fan, J.P., Luo, H.Z., and Satoh S.I. 2008. A novel approach for filtering junk images from Google search results. In Proc. of MMM2008, Vol. Part II, 1-12.

[0031] [15] Yuejie ZHANG, Yi ZHANG, Shuai REN, Cheng JIN, Xuanjing HUANG. 2013. Junk Image Filtering via Multimodal Clustering for Tag-based Social Image Search, Vol. 9 (6): 2415- 2422.

[0032] [16] Lowe, D. G. 2004. Distinctive image features from scale-invariant keypoints. International journal of computer vision, 60(2), 91-110.

[0033] [17]Bay, H., Tuytelaars, T., & Van Gool, L. 2006. Surf: Speeded up robust features. In Computer Vision-ECCV 2006 (pp. 404-417). Springer Berlin Heidelberg.

[0034] [18]Csurka, G., Dance, C., Fan, L., Willamowski, J., & Bray, C. 2004. Visual categorization with bags of keypoints. In Workshop on statistical learning in computer vision, ECCV Vol. 1, p. 22.

[0035] [19]Elkan, C. 2003. Using the triangle inequality to accelerate k-means. In MACHINE LEARNING-INTERNATIONAL WORKSHOP THEN CONFERENCE- Vol. 20, No. 1.

[0036] [20]Koheri Arai and Ali Ridho Barakbah. 2007. ″Hierarchical K-means: an algorithm for Centroids initialization for k-means,″ department of information science and Electrical Engineering Politechnique in Surabaya, Faculty of Science and Engineering, Saga University, Vol. 36, No.1.

[0037] [21] Greg Hamerly. 2010. ″Making k-means even faster″, In SIAM International Conference on Data Mining.

[0038] [22] Drake, Jonathan, and Greg Hamerly. 2012. ″Accelerated k-means with adaptive distance bounds.″ 5th NIPS Workshop on Optimization for Machine Learning.

[0039] [23] Lazebnik, S., Schmid, C., Ponce, J. 2006. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories.

[0040] [24]MarcAurelio Ranzato, F., Boureau, Y., LeCun, Y. 2007. Unsupervised learning of invariant feature hierarchies with applications to object recognition. In: Proc. Computer Vision and Pattern Recognition Conference CVPR07.

[0041] [25]Serre, T., Wolf, L., Poggio, T. 2005. Object recognition with features inspired by visual cortex. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Volume 2, 994.

[0042] [26] Zhang, H., Berg, A., Maire, M., Malik, J. 2006. SVM-KNN: Discriminative nearest neighbor classification for visual category recognition. In: Proc. CVPR. Volume 2, 2126-2136.

[0043] [27] Makadia, A., Pavlovic, V., Kumar, S. 2008. A new baseline for image annotation. In: Proc. ECCV, 316-329.

[0044] [28] Torralba, A., Fergus, R., Weiss, Y. 2008. Small codes and large image databases for recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, 2008. CVPR 2008, 1-8.

[0045] [29] Bosch, A., Zisserman, A., Munoz, X. 2007.Representing shape with a spatial pyramid kernel. In: Proceedings of the 6th ACM international conference on Image and video retrieval, ACM, 408.

[0046] [30] Silla Jr, C. N., & Freitas, A. A. 2011. A survey of hierarchical classification across different application domains. Data Mining and Knowledge Discovery, 22(1-2), 31-72.

[0047] [31] Yuan, X., Lai, W., Mei, T., Hua, X. S., Wu, X. Q., & Li, S. 2006. Automatic video genre categorization using hierarchical SVM. In Image Processing, 2006 IEEE International Conference on (pp. 2905-2908). IEEE.

[0048] [32] Cesa-Bianchi, N., Gentile, C., & Zaniboni, L. 2006. Hierarchical classification: combining Bayes with SVM. In Proceedings of the 23rd international conference on Machine learning (pp. 177-184). ACM.。

发明内容

[0049] 本发明的目的在于提出一种基于网购平台的商品图像类别预测方法，从而提升用户在使用网购平台时的体验。

[0050] 为此，本发明基于当前网购平台上大规模的商品图像数据，利用计算机科学中多媒体信息处理、人工智能和机器学习等技术，为实现商品图像类别的预测构建了一套完整的框架。这套框架可以自动地分析用户所输入的图像，利用由海量训练数据所产生的分类模型，预测输入图像在网购平台中可能的类别信息，并将其向用户返回，便于用户检索和浏览与输入图像相关的产品。因此，要实现商品图像类别的预测，需要包含以下步骤：

[0051] (1) 训练图像的获取——向当前的网购平台爬取商品图像和图像相关标注，并初

步清理垃圾数据，为训练图像分类模型提供数据；

[0052] （2）图像特征提取——选择特定的特征表达方法，将爬取的图像从点阵表示转化为特征表示；

[0053] （3）不相关图像过滤——利用2中所产生的特征表达，清除与标注不相关的商品图像；

[0054] （4）图像特征训练——对于图像的兴趣点特征表达，进一步训练BOW词典，将图像转化为词包表达；

[0055] （5）多层次图像分类——根据商品图像的词包表达，利用图像的BOW特征，训练多层次的图像分类模型，并应用于用户输入图像的类别预测；

[0056] （6）相关图像选择——根据(5)中所提供的类别预测，选取返回给用户的相关图像。

[0057] 其中，图像特征提取、不相关图像过滤、图像特征训练、多层次图像分类则是本发明的核心部分。

[0058] 附图1为本发明系统框架图，包括训练图像的获取、图像特征提取、不相关图像过滤、图像特征训练、多层次图像分类、相关图像选择六个模块。其中，训练图像获取模块从网购平台获取商品图像数据，图像特征提取模块提取商品图像特征供图像特征训练和不相关图像过滤模块使用，过滤不相关图像后，利用训练完成的特征使用多层次图像分类模块对用户输入图像的类别进行预测，最后利用相关图像选择模块将相关图像向用户返回。

[0059] 本发明的关键点是四个核心模块的算法包括六个模块的商品图像类别预测应用框架。四个核心模块算法是：(1) 图像特征提取和相关性定义算法；(2) 不相关图像过滤算法；(3) 图像特征训练算法；(4) 多层次图像分类算法。利用以上四个核心算法以及辅助这些算法运行的两个模块设计一种基于网购平台的技术框架。

[0060] 下面详细介绍本发明基于网购平台的商品图像类别预测方法及组成该框架的四个核心算法和两个附属模块：

[0061] 系统框架

[0062] 该框架可分为 (1) 图像特征提取；(2) 不相关图像过滤；(3) 大规模图像特征训练；(4) 多层次图像分类等四个核心模块和训练图像获取、相关图像选择两个附属模块：此外，在该应用框架的各个模块中，还将运用一些目前已经比较成熟的算法和技术。

[0063] (1)图像特征提取模块：在互联网中，大多数图像是以位图的方式存储在jpeg、png、gif等图像格式中。这种以点阵方式保存的图像，具有表述简单，方便压缩等特点。但是，在使用计算机视觉的方法对数字图像进行处理和分析时，这种表述方法的图像往往不能直接使用，而需要将图像转化为其他更接近人对图像认知的方法进行重新表述。这种重新表述的过程就是图像特征的提取。在特征提取的过程中，可以根据不同的需要从不同的角度对图像进行表述，这些表述可以是图像的明暗、颜色、纹理、兴趣点等。为了将提取的图像特征应用于后续不相关图像过滤以及图像分类过程中，图像特征提取模块不仅仅要定义图像的特征，同时也需要定义不同图像间在某个特定特征上的相关性。这种特征表达相关性的定义，可以为图像间在特征空间上的相似度计算奠定基础。

[0064] (2)不相关图像过滤模块：网购平台上的商品图像都是由商户上传并标注的，这种由社会化用户上传的标注总是存在着与实际图像并不完全符合的问题。这种问题的产生存

在多方面的原因，如网购平台商品类别不健全、上传者和浏览者之间语义鸿沟以及上传者对搜索引擎的过度优化等。若训练数据中存在大量标签与图像间不正确的匹配，训练产生的分类模型将会应噪声过大而没有意义。因此，在将直接从网购平台中爬取的商品图像及相关标注作为训练数据前，需要对商品图像中不相关的标签做一次清理工作。将具有更大相关性的数据和其标注，作为训练数据保留下来。这项工作从另一个角度看，即过滤相同标签下与标签不相关的图像。

[0065] (3)图像特征训练模块：根据当前流行的BOW分类模型，图像最终需要表达成词包的形式。词包本身则是由图像中每一个视觉词出现的频率所组成。而视觉词则源于视觉词典，是由训练样本训练所产生的。在网购平台商品图像类别预测的应用中，由于每一幅图像中都能够抽取数百个与尺度、大小、旋转无关的兴趣点，因此，相比图像的数量，视觉兴趣点的数量更加惊人。而要将这些视觉兴趣点训练成视觉词典，则需要有支持大规模数据的聚类算法实现。具体的，在本发明中，选取了相比其他聚类运行效率更高的K-means算法作为基础，并且在K-means算法上作进一步优化，以实现大规模图像特征的训练，最终实现图像的视觉词包表达。

[0066] (4)多层次图像分类模块：商品图像在网购平台中的特点除了数量巨大以外，类别也特别多。普通的分类方法往往专注于解决两类或者少量类别的分类问题。而在商品图像类别预测任务中，直接应用这些分类模型往往会产生分类效果急剧下降和时间复杂度迅速增长的问题。比如，其中一些相对分类效果较好的方法，会随着类别数量的增长而使分类模型的训练时间和利用分类模型预测新样本的时间成平方级地增长。这在不但图像数量巨大，类别数量也巨大的商品图像类别预测中是不适用的。幸运的是，在网购平台中，商品的类别总是以层次结构呈现，利用这种人为定义的层次结构，可以将商品图像的分类过程层次化地进行。这样不仅能够加快训练和预测的速度，如果针对不同类别的商品训练不同的模型，还能够提升商品预测的准确率。同时，这种层次化的分类模型训练方式，也更易于保持训练分类模型时正负样本的平衡性。

[0067] (5)训练图像获取模块和相关图像选择模块：由于本发明所使用的方法需要网购平台上的商品图像及其标注信息数据所支撑，所以需要向网购平台爬取海量的训练图像。然而，为了有效地利用网购平台上的商品图像数据，使用科学的方法对网购平台上的商品图像及其标注进行采样至关重要。这是训练图像获取模块的主要工作。另一方面，在通过商品图像类别预测系统对商品图像的类别预测以后，将相关的商品图像直接返回给用户能够极大的提升用户对于系统使用的体验，因此，自动选择部分类别相关的图像返回给用户，也是系统框架中所需要的模块之一。

[0068] 下面对各部分的具体内容作进一步描述。

[0069] 图像特征提取

[0070] 本节内容所述的特征提取只包含图像底层特征的提取，而并不包含词包特征。词包特征将在下文大规模图像特征训练节详细描述。为了能够全面地描述图像各方面的特点，必须从多个角度选取图像的视觉特征。因此，本发明从颜色、纹理和兴趣点三个角度出发，分别为每个方面选取一种适应于商品图像的特征。这三种图像的视觉特征分别是：(1)基于网格的颜色直方图特征；(2) Gabor纹理特征；(3) SURF(Speeded Up Robust Feature)兴趣点特征。

[0071] 图像的颜色特征是人类对图像认知中最直观的特征。实际上计算机中彩色图像的点阵表达也是通过对于描述图像每一个像素的颜色完成的。然而，目前常用的bmp、jpeg、gif和png等图像格式一般都遵循红绿蓝(Red Green Blue, RGB)的颜色空间，这与人类对色彩的认知并不相同。因此，为获取更符合人类认知的颜色特征，本发明先将图像从RGB空间转换为色彩饱和度明暗(Hue Saturation Value, HSV)空间。

[0072] 颜色直方图是描述图像颜色特征的重要方法，这种方法统计每种颜色在单幅图像中出现的概率，并将所有颜色的出现概率组成向量。然而，直接使用这种简单的方法对图像颜色特征进行描述有两个明显的问题：(1) 按照当前流行的图像编码方式，在RGB空间中每个通道均有8bit用于表示该通道的值，因此共有2^{24}种颜色，若要按照此方法描述一幅图像，则向量将有2^{24}维，这是在当前技术条件下无法接受的；(2) 图像自身的颜色直方图只能表达图像全局的颜色特点，即对于出现在不同位置的相同色块无法区分。为克服问题(1)，可将颜色空间划分为多个区域，将同一个区域中的所有颜色看做同一种颜色，而这些区域则被称为桶。然而，这种处理方式在大幅度减少颜色数量的同时，也会使得问题(2)更加突出。本发明选用较为常见的36个桶的方式。为克服问题(2)，可以将原始图像划分为多个网格，利用不同数量的网格划分方式，实现不同尺度的颜色特征的表达。考虑到本发明应用于商品图像，商品图像本身往往只描述少量物品，且物品一般均位于图像的中心位置，因此图像的局部颜色特征并不如普通图像重要。因此，本发明仅选取3种尺度的网格用于描述图像颜色特征。每种尺度的划分均是对上一种尺度中每一个网格进行田字划分。共$\sum_{r=0}^{2} 2^r \times 2^r = 21$个网格，$21 \times 36 = 756$维颜色直方图特征。附图2描述了选取4种尺度时的网格划分方式。

[0073] 基于这些颜色直方图特征，两幅图像u和v之间颜色直方图的相似性$k_c(u,v)$可以定义为：

[0074] $$\kappa_c(u,v) = \frac{1}{\sum_{r=0}^{R-1}\frac{1}{2^{R-r+1}}}\sum_{r=0}^{R-1}\frac{1}{2^{R-r+1}}N_r(u,v) \qquad (1)$$

[0075] 其中，R=3，是所取网格划分图像尺度的种类数量；$N_0(u,v)$表示原始分辨率的颜色直方图相似性；$N_r(u,v)$表示第r种分辨率的颜色直方图相似度。

[0076] $$N_r(u,v) = \sum_{i=1}^{36} Norm\left(H_i^r(u), H_i^r(v)\right) \qquad (2)$$

[0077] 其中，$H_i^r(u)$和$H_i^r(v)$分别表示图像u和v中，第r种分辨率网格划分中第i 个格子的颜色直方图相似度；Norm代表的是二阶标准距离。

[0078] 除颜色特征以外，纹理特征也是图像重要的传统特征。与颜色特征相同纹理特征在不同尺度的表现也不同。另外纹理特征还具有方向性，因此本发明选用三个尺度四个方向共12个21×21像素点的Gabor滤波器构造图像纹理特征。将原始图像转化为灰度图像后，分别使用这些滤波器对图像做滤波操作。分别计算滤波后12幅图像所有像素点的均值和方差，可以得到$12 \times 2 = 24$维的Gabor纹理特征。

[0079] 基于上述纹理特征，两幅图像u和v之间Gabor纹理的相似性$\kappa_t(u,v)$可以被定义

为：

[0080] $$\kappa_t(u,v) = e^{-d_t(u,v)/\sigma_t}$$

[0081] $$d_t(u,v) = Norm\left(g_i(u), g_j(v)\right) \quad (3)$$

[0082] 其中，σ_t代表所有图像$d_t(u,v)$的均值；$g_i(u)$和$g_j(v)$分别代表图像u的第i个和图像v的第j个Gabor描述子(包括均值和标准差)。

[0083] 图像的颜色特征和纹理特征尽管已经经过尺度上的处理，但其本质上依然是全局特征。因此为了更全面地描述图像，本发明引入兴趣点特征作为局部特征。SIFT算法和SURF算法是两种经典的兴趣点提取算法。考虑到训练数据规模巨大，本发明选用执行更快，表达也更简单的SURF算法。由于不同图像中的兴趣点数量并不相同，所以每幅图像的SURF特征数量并不固定。但是SURF算法提取的每个兴趣点有64维。

[0084] 基于上述SURF算法，由于不同图像间兴趣点的数量不同，因此难以直接计算两幅图像间基于兴趣点特征的相似性。为此，本发明首先将两幅图像的兴趣点做一一配对(兴趣点数量多的图像有部分兴趣点没有配对)，使得所有兴趣点配对间二阶标准距离之和最小。该配对可以用二分图匹配算法实现。至此两幅图像u和v之间SURF特征的相似性$\kappa_s(u,v)$可以被定义为：

[0085] $$\kappa_s(u,v) = e^{-d_s(u,v)/\sigma_s}$$

[0086] $$d_s(u,v) = \frac{\sum_i Norm(s_i(u), s_i(v))}{\sum i} \quad (4)$$

[0087] 其中，σ_s代表所有图像$d_s(u,v)$的均值；$s_i(u)$和$s_i(v)$分别代表图像u的第i个兴趣点和图像v中与其配对的兴趣点。

[0088] 最后，视觉相似性可以通过一个混合的线性加权统计出最终的结果，因此图像u和v之间的视觉相似性可以定义为：

[0089] $$\kappa(u,v) = \beta_1\kappa_c(u,v) + \beta_2\kappa_s(u,v) + \beta_3\kappa_t(u,v)$$

[0090] $$\sum_{i=1}^{3}\beta_i = 1 \quad (5)$$

[0091] 其中，β_i表示每一种特征所占的权重，根据所有图像间$\kappa_c(u,v)$、$\kappa_s(u,v)$和$\kappa_t(u,v)$的方差分配。将所有的特征的相似性合并起来可以将图像间的相似性问题简化，使后续应用更易处理。

[0092] 不相关图像过滤

[0093] 基于图像两两间视觉相似度的定义，可以将图像及图像间的关系建立带权的无向图模型。其中，每一幅图像都成为图中的一个点，图像两两间的相似度则成为连接两点间边的权重。这样，由图像两两间相似性组成的相似性矩阵就是其按照上述规则所建立图模型的邻接矩阵。

[0094] 对于大规模社会化标注的图像，使用有监督的方法对不相关的图像进行过滤往往需要利用人工重新标注大量信息。这类方法虽然效果较好，但是在类别数量巨大的商品图

像面前，需要大量的人力资源，所以并不适用。因此本发明选用了无需人工重新标注的无监督的方法。

[0095] 考虑到社会化用户为商品图像标注的类别标签在许多情况下都是准确的情况，可以认为，在具备同一类别标签的所有商品图像中，具有大量的图像是与该标签是相关的。进一步而言，对于属于相同类别的商品图像，在视觉特征上具有相关性。另一方面，对于与标签不相关的商品图像，往往会属于多个不同的类别，这些图像在视觉特征上不仅与那些相关图像相似性较小，互相之间的视觉特征差距也较大。因此，若能将所有图像聚类成一类内部相似性很大，而该类与其他图像的类间相似性很小，则可以对不相关图像作一定程度上的过滤。

[0096] 要通过上述方法对图像聚类，首先需要定义图像类与图像类之间的类间距离以及单个图像类内部的类内距离。对于两个图像类G_i和G_j而言，它们的类间距离可以定义为：

[0097] $$s(G_i,G_j)=\frac{\sum_{u\in G_i}\sum_{v\in G_j}\kappa(u,v)}{\sum_{u\in G_i}1\times\sum_{v\in G_j}1} \tag{6}$$

[0098] 而对于图像类G_i而言，其类内距离可以相应的定义为：

[0099] $$s(G_i,G_i)=\frac{\sum_{u\in G_i}\sum_{v\in i(u\neq v)}\kappa(u,v)}{\sum_{u\in G_i}1\times\sum_{v\in G_i}1} \tag{7}$$

[0100] 对于类内距离大于所有与其他类的类间距离的图像类，应当将其再度划分；对于两类类间距离小于两类类内距离的，则应当将两类合并。通过这样两条规则，结合标准割算法(Normal Cut, Ncut)，可以对图像分类进行不断迭代。当迭代次数达到一定值或类别数量达到预设最大类别数量K时，停止迭代。这时，可从结果中选取其中图像数量最多的类，将类中的所有图像作为相关图像，而将其他类别的图像作为不相关图像。虽然在该方法所得到的结果中，作为不相关图像的类别内依然会存在大量的相关图像，但作为相关图像的类别里，图像间的视觉相似性更大，与标签相关的可能性更高。对于可以利用海量的商品图像的应用而言，在过滤不相关图像的过程中，流失少量相关图像也是可以接受的，只要保证被排除的相关图像与不相关图像比例比原本的相关图像与不相关图像的比例更小。这样，对于所有图像使用图模型上的分裂合并算法后，选取其中最大的类别，即可实现不相关图像过滤，如附图3所示。

[0101] 图像特征训练

[0102] 为使用BOW特征训练商品图像的分类模型。首先需要对商品图像抽取兴趣点特征。在本发明中，考虑到应用需要使用海量商品图像数据的特点，选用SURF算法作为提取图像兴趣点特征的算法。相比经典的特征点提取算法SIFT，SURF算法不仅在特征点提取时具备更高的效率，而且最终对于兴趣点的特征表达也仅仅需要64维，只有SIFT算法128维的一半。这能从理论上为BOW词典训练工作提升一倍效率。

[0103] 网购平台商品图像的类别预测任务介于图像分类与图像检索之间，在部分图像上具备图像检索的特性，而部分图像又体现图像分类的特性。因此本发明选取16384作为BOW词典的规模，该规模大于一般图像分类应用而小于图像检索应用所使用的词典。

[0104] 在目前的网购平台中，详细的商品类目有数万之巨，即使是基本商品的类目，也有

数百。在这样的背景下，即使只判别商品的基本类目，每类商品选取数千幅图像作为训练分类模型使用，也需要有百万级别的商品图像。在使用SURF算法对商品图像抽取兴趣点的过程中，平均每幅图像会被抽取数百个兴趣点。因此，用于训练BOW模型词典的兴趣点数量就至少有数亿的规模。即使是将所有兴趣点的64维SURF特征存入运行系统的内存中，内存的占用也将达到近百G的规模。在常用的聚类算法中，以ap-clustering为代表的基于样本点间邻接矩阵的聚类算法在这样的样本规模下，所需要的空间将会达到目前大规模集群也难以处理的百PB级别，运算量则更是远在此之上。因此，BOW词典的训练算法，只能局限于无需计算样本点间邻接矩阵的算法之内。在无须计算样本点邻接矩阵中的聚类算法中，最为著名的是K-means算法，该算法不但应用广泛，运算速度相对较快，且随着迭代执行的运行，聚类效果会逐渐收敛至最佳。这种算法的优势在于，即使其收敛的过程需要执行数千轮迭代，只要经过几十轮的迭代，就能够得到接近最终迭代收敛结果的一个解。

[0105] 然而，K-means算法也有着巨大的缺陷，就是其算法最终结果收敛的效果很大程度上依赖于初始中心的选择。在小规模数据中，K-means算法往往会被多次运行，而每次运行都会选择不同的随机初始中心，最后选择多次运行的最佳结果作为最终结果。这种方法在样本点和中心点数量较少时可以有较大可能得到全局较优的初始点分布，但当样本点数量和中心点数量增加时，每一个初始中心点都处于较优位置的可能性成几何级数下降。因此，这种方法在面对大规模数据时，并没有太大的实用性。另外一些基于规则的初始点选择方法则与数据规模大小关系并不大。例如最大最小距离算法是每次选择一个能使与当前所有的最小距离最大化的样本点作为一个新的中心，直至得到所有初始中心点。但这种方法一方面由于规则本身限制随机性较小，另一方面，在最大最小距离时，所需要的运算开销，也远比K-means算法本身更大而与需要计算邻接矩阵的聚类算法类似。因此在大规模数据的条件下也无法使用。

[0106] 相较上述初始点选择方法而言，一种利用层次聚类和K-means本身结合的初始点选择方法则能够满足在海量数据条件下的诸多限制而成为本发明所使用的初始点选择方法。这种被称为层次K-means聚类(Hierarchical K-means)的算法其具体的算法流程如下：

[0107] (1)设置迭代次数$i = 0$；

[0108] (2)利用K-Means算法对原始数据进行聚类，K个随机点作为原始聚类中心，算法达到收敛条件后得到K个聚类中心$\{C_k^i\}$；

[0109] (3)当i<M(M为预设最大迭代次数)时，i=i+1，重复执行(2)；

[0110] (4)将得到的所有M * K个聚类中心作为样本点($g_j = C_k^i, j = i * K + k$)，执行基于重心距离的层次聚类算法(Centroid-Linkage Hierarchical Clustering)

[0111] a)该算法将所有初始样本看作类中心$C_j = g_j$(1<=j<=N * K)

[0112] b)计算任意两类中心之间的距离作为类与类的距离，将距离最小的两类合并：

[0113] $\delta(C_S, C_T) = \mathrm{d}(\mathrm{cs}, \mathrm{ct})$

[0114] $$\mathrm{vs} = \frac{1}{|C_S|}\sum g_j \in C_S, \quad vt = \frac{1}{|C_T|}\sum g_j \in C_T \tag{8}$$

[0115] c)重复执行b)直至最终只剩下K类；

[0116] (5)以(4)中得到的K个类中心作为初始类中心，执行按照步骤(1)-(3)K-means算

法，直至收敛。得到最终的K个类中心。

[0117] 这种方法实际上利用多次随机初始中心的K-means算法本身，将其执行的结果作为层次聚类算法的样本点。当对这些样本点完成层次聚类之后，层次聚类的结果能在一定程度上表现原有样本点的疏密程度，并且能避免在随机选择初始点方法中有较大几率选到距离接近的点作为初始点的情况。而其代价，则和多次随机初始中心点执行K-means算法的方法一样，需要多次重复执行K-means算法。但根据不同的初始中心点执行K-means的任务可以轻而易举地划分到多个运算单元中执行，因此该方法在该层面的并行性良好。

[0118] 但是，在当前网购平台商品图像的规模下，K-means算法本身的计算量也相当惊人。朴素的K-means算法的计算复杂度是中心点数量K、预设最大迭代次数M、样本点数量N以及样本维度D的乘积。按照本节开头所述的规模，单纯其计算样本点与中心间欧几里得距离所需要用到的计算量就达到数十PB。在完美并行的条件下也需要有包含数百台计算机的集群才能在短时间内运算完成。为此本发明提出一种能够保证结果与朴素K-means算法一样，但效率提升数百倍的加速算法。

[0119] 这种算法的大体思路是利用每个样本点与其上一轮所分配中心的距离和三角形不等式模型，推测其与本轮所有中心的距离关系，从而大幅减少计算该样本点与本轮所有中心点距离的运算次数。为描述其具体算法，首先定义相关变量如下：

[0120]

x_i	第 i 个样本点的向量表示
a_i	第 i 个样本点当前所属中心的编号
$disx_i$	第 i 个样本到其当前所属中心的距离
c_j	当前一轮待分配的第 j 个中心的向量表示
c_j'	当前一轮待分配的第 j 个中心的向量表示
$mindisc_j$	当前一轮待分配的第 j 个中心与其最近中心的距离
$mindisc_j'$	上一轮第 j 个待分配中心和当前待分配中心的最小距离
$secmindisc_j'$	上一轮第 j 个待分配中心和当前待分配中心的第二小距离
$disc_j$	第 j 个中心向量这一轮和上一轮的距离
u_i	当前一轮第 i 个样本点到其最近中心的上界
l(i,j)	当前一轮第 i 个样本点到第 j 个待分配中心距离的下界
d(a,b)	a，b 连个向量的距离

[0121] 根据上述定义，如参考文献[19]中所述，可以利用3个三角形不等式优化相关的距离计算，其迭代过程中判断样本点所属中心的关键执行步骤如下：

[0122] (1)若 $2*\left(disc_{a_i}+disx_i\right)<mindisc_i$ 成立，则第i个样本点直接分配给中心 a_i，否

则$u_i = d(x_i, c_{a_i})$；

[0123] (2)若$2 * u_i = d(c_{a_i}, c_j)$成立，则第i个样本点至中心$a_i$的距离小于其与第j个中心点的距离，可省去其与第j个中心点的距离计算；

[0124] (3)若$l(i,j) = d(x_i, c_j') - disc_j, l(i,j) > u_i$成立，则第i个样本点至第j个中心的距离大于其与中心a_i的距离，可以省去其与第j个中心点的距离计算。在步骤(2)，(3)均不满足的条件下，需要计算第i个样本点与第j个中心点的距离，更新u_i。

[0125] 根据上述的步骤K-means算法在K较大的数据集上，相比朴素K-means算法均能得到上百倍的加速。然而，该算法需要建立每个样本点到每个待分配中心距离下界的表，该表的规模是样本点数量N和中心点数量的乘积，在本节所述数据量条件下无法存放于内存之中，因此该表将严重影响算法在大规模数据下效率。

[0126] 为此，在本发明算法中，可以将该条加速优化删去，并增加一个新的步骤：

[0127] (4)若$2 * disx_i + mindisc_j' < secmindisc_j'$成立，则第i个样本i直接分配给中心$a_i$，否则执行步骤(1)。

[0128] 步骤(4)在步骤(1)之前执行，该步骤与步骤(1)相似，但能从另一个角度发挥作用，因此在没有规则(3)的情况下，是规则(1)的一个良好补充，能够对K-means算法产生加速效果。另一方面，由于每次样本点的分配操作只需要用到待分配中心、样本点本身以及少量临时数据，因此该步骤具有良好的并行性，在集群中，可以实时分配到多个计算节点进行运算而不会受限于单台计算机。

[0129] 至此，利用海量商品图像的SURF特征训练BOW词典的算法已经完成，该算法在本节所述的数据规模下，能够在多个小型集群中快速计算完成，并且具有较好数据扩展性和并行性。

[0130] 多层次图像分类

[0131] 在完成BOW的词典训练后，为实现图像分类模型的训练，首先需要对将图像从特征点表示转化为词包表示。本发明对于特征点的处理采用选择词典中与其欧几里得距离最近的词作为该特征点的表达。将所有特征点转化为视觉词后，对每一幅图像统计所有词出现的频率，将其作为词的BOW模型特征。每幅商品图像的特征，根据词典大小，是一个16384维的向量。

[0132] 本发明使用基于SVM分类方法的算法训练分类模型。在分类问题中，SVM具有广泛的适用性，并且不同的核函数具备不同的效果。RBF核作为应用最多的SVM核函数，在大多数应用背景下具有较好的效果。其衡量两个向量x与z之间距离的定义为：

[0133] $$K(x,z) = exp\left(-\frac{\|x-z\|^2}{2\sigma^2}\right) \quad (9)$$

[0134] 在计算RBF核函数值的过程中，对于向量x和z的对应位置i存在三种现象：

[0135] (1) $x_i = 0, z_i = 0$；

[0136] (2) $x_i = 0, z_i > 0$ 或 $x_i > 0, z_i = 0$；

[0137] (3) $x_i > 0, z_i > 0$。

[0138] 由于词典大小为16384，而每幅图片包含词的数量只有数百，图像的BOW特征向量

是稀疏的。并且，现象(1)占了绝大多数，而在剩下的可能中，现象(2)也比现象(3)更多。假设两个向量拥有的非零向量不同，但非零向量的数量大小相似，则在使用RBF核函数计算两个向量之间的距离时，由于上述情况，结果会倾向于被第二种现象所产生的值主导。对于向量而言，所有现象(2)的结果值取决于两个特征向量自身的性质，两个向量间的相互关系对其影响不大。另一方面，由于RBF核函数对于每一对相同维度的值都是采用平方的方式计算距离，其特征向量自身的特征被进一步放大。更极端的情况是，部分图像中，某些视觉词出现的次数是其他词的十多倍，经过平方放大后有百倍以上的影响，这对于衡量特征向量间距离是极为不利的。

[0139] 事实上，在稀疏向量的距离的计算中，重要的是上述现象(3)中所表现的情况。所有现象(3)的情况，直接描绘两个特征向量间的关系。由于出现的次数少，而被大量现象(2)所计算得到的值所掩盖，极大地影响了SVM分类模型的效果。因此，本发明使用一种改进的RBF核——X^2-RBF核作为SVM核函数。该核函数的定义为：

[0140] $$K(x,z) = exp\left(-\frac{(x-z)^2}{2\sigma^2(x+z)}\right) \qquad (10)$$

[0141] 从公式(10)中可以看到，X^2-RBF核在处理现象(2)时，相比原始RBF核去除了平方放大操作，而仅仅将值累加。而且，这种将所有现象(2)直接累加的结果就是两幅图像所包含的不同视觉词所占比例之和。对于现象(3)而言，X^2-RBF核也对其做了一定的调整，由于现象(3)的情况较少发生，所以值相对较小，因此按照其所占比例做一定程度的放大，能够将其影响扩大到应有的程度。在使用实际数据的实验中，对于SVM的核函数做这样的调整，效果是十分明显的。

[0142] SVM是一个面向两类问题的分类器。要将SVM用于多类分类的问题，常用的方法主要有两种。

[0143] (1)一对多算法(one-versus-rest, 1-v-r SVMs)——该方法一次用一个两类SVM分类器将每一类与其他所有类别区分开来得到k个分类模型。分类时讲位置样本分类为具有最大分类函数值的那类。

[0144] (2)一对一算法(one-versus-one, 1-v-1 SVMs)——该方法在每两类间训练一个分类器，因此对于一个k类问题，将有k(k-1)/2个分类模型。当对一个未知样本进行分了时，每个分类器都对其类别进行判断，并为相应的类别投票，最后得票数最高的类别作为该未知样本的类别。

[0145] 这两种SVM的方法各有优缺点。对于k类分类问题，一对多算法只需要训练k个分类模型，在预测时也只需要使用k个分类模型对未知样本进行预测。但这种一类与其他所有类别区分开作为正负样本的方式，在k的数量较大时，正负样本的数量极不平衡。这种负样本是正样本数量几百倍的情况将极大地影响SVM分类器的分类效果。而对于一对一的算法，虽然每次训练都只使用两个类，正负样本的数量能够很容易达到平衡，但其训练过程需要有$k(k-1)/2$个分类模型，随着类别数量k的增长，训练的时间将会呈平方的关系增长。另一方面，即使训练模型可以通过高性能的集群离线完成，一对多的算法在预测一个未知样本时也需要使用全部$k(k-1)/2$个分类模型，并统计所有分类模型得到的结果才能做出最终的分类判断，这在类别数量巨大的商品图像分类问题中也是难以承受的。

[0146] 幸运的是，根据人类对世间万物认知的习惯，商品本身的类别具有层次性。这种层次性不但能够帮助用户更好地检索商品，同时属于相同类别中的商品，也具有一定的相似性。这样，利用商品类别本身的层次属性，可以从商品类别的最高层开始，自上而下地对商品的类别进行预测。这种层次分类的方法将商品的分类关系表达成树或者有向无环图(Direct Acyclic Graph, DAG)的结构，当树的节点具有多个子节点或DAG的节点具有多个出度时训练一个多类分类问题的模型。其中，DAG与树结构的区别在于，用DAG表示的类别层次结构更接近于现实的情况，可以容许一个类别从属于多个父类别的情况，而树结构则对于每个节点只能拥有一个父节点，如附图4所示。但这种情况会增加该类别被分类到的可能性，因此本发明选用树结构表示类别的层次结构。这种树结构的关系可以用数学语言定义为：

[0147] $\forall c_i, c_j \in C$，如果 $c_i \prec c_j$ 那么 $c_j \nprec c_i$

[0148] $$\forall c_i \in C, c_i \nprec c_i \tag{11}$$

[0149] $\forall c_i, c_j, c_l \in C$，如果 $c_i \prec c_j$ 并且 $c_j \prec c_l$ 可以推出 $c_i \prec c_l$

[0150] 其中，c_i, c_j, c_l 分别代表第i，j，l个类别，C表示所有类别的集合。

[0151] 在这样的树结构中，存在多种类别划分的策略用于层次分类，在介绍具体的分类方法之前，本发明对相关的变量作如下定义：

[0152]

Tr	所有训练样本
$Tr^+(c_j)$	对于类别 c_j 而言的所有正样本
$Tr^-(c_j)$	对于类别 c_j 而言的所有负样本
$\uparrow(c_j)$	c_j 的父类别
$\downarrow(c_j)$	c_j 的所有子类别集合
$\Uparrow(c_j)$	c_j 的所有祖先类别集合
$\Downarrow(c_j)$	c_j 的所有子孙类别集合
$\leftrightarrow(c_j)$	c_j 的所有兄弟类别集合
$*(c_j)$	c_j 中所有样本点的集合

[0153] 基于相应的定义，可以有相应的五种解决层次分类问题正负样本定义的方法：

[0154] (1) $Tr^+(c_j) = *(c_j), Tr^-(c_j) = Tr \backslash *(c_j)$

[0155] (2) $Tr^+(c_j) = *(c_j), Tr^-(c_j) = Tr \backslash *(c_j) \cup \Downarrow(c_j)$

[0156] (3) $Tr^+(c_j) = *(c_j) \cup \Downarrow(c_j), Tr^-(c_j) = Tr \backslash *(c_j) \cup \Downarrow(c_j)$

[0157] (4) $Tr^+(c_j) = *(c_j) \cup \Downarrow(c_j), Tr^-(c_j) = Tr \backslash *(c_j) \cup \Downarrow(c_j) \cup \Uparrow(c_j)$

[0158] (5) $Tr^+(c_j) = *(c_j) \cup \Downarrow(c_j), Tr^-(c_j) = \leftrightarrow(c_j) \cup \Downarrow$

[0159] 而在商品图像类别预测中，实际上只有最底层的叶子节点的类别才包含相应的样本点，具有实际意义，其他类别均为虚拟类别。因此在这五种层次分类的正负样本定义方法

中，(1)、(2)、(5)的定义将所有类别作为最终的类别划分，与相应的子类别存在互斥关系，因此这三种正负样本的定义方式并不适合。在(3)的定义中，负样本包含所有除c_j子孙类别中样本外的所有样本，这对于我们最终要解决的叶子节点的分类问题而言，又回到没有使用层次分类模型的情况。因此在本发明中使用(4)中对正负样本的定义。这种定义方式可以通过自顶向下的顺序，对叶子节点类别进行分类模型训练。每次分类模型的训练只包含同一父亲节点的所有兄弟节点。在实际的商品类别预测的情况中，兄弟节点的数量往往在数十个。这种情况可以很好的通过普通的多类SVM模型解决。为达到更好的效果，本发明选取相对一对多算法精度更高的一对一的算法，解决该小规模的多类分类问题。经过自顶向下，3-4次小规模多类别的分类之后，即可得到样本的最终类别预测。

[0160] 然而，这样每一轮都严格分配一个类别层次分类方法有一种致命的缺陷，即当高层的分类发生错误时，低层的分类将会完全没有意义。并且高层次的类别由于包含了大量子类别，其在视觉表现上十分复杂。这种特点将导致分类模型分类性能的下降。为解决这个问题，本发明在上述层次分类的基础上，加入一些潜在可能分类，从而使高层误分类情况能够得到缓解。其具体步骤如下：

[0161] (1)在最高层的类别中，根据一对一算法预测时的排序结果，选择前五个类别作为商品图像备选的类别

[0162] (2)分别将商品图像应用于上一步所产生的五个类别中，亦根据一对一算法每类产生五个子类别，得到25个相对于上一步中孙子代的备选类别

[0163] (3)为步骤(2)中的25个类别训练一对一的多类SVM分类模型，根据其投票机制，选取排名前五的类别循环执行步骤(2)，直至所得到的五个类别均为叶子类别。

[0164] 至此，本发明基于BOW特征的图像层次分类模型已经完成。该分类模型能够为待分类的样本商品图像提供五个备选类别可能，并能为这五个备选类别排序。需要注意的是，为保证这种层次分类方法的效果，商品最终类别在类别树中的深度应当一致，不同兄弟类别之间的训练样本数量也应该尽可能接近。

[0165] 训练图像获取和相关图像选择

[0166] 为实现商品图像类别的自动化预测，需要从网购平台获取的商品图像作为训练图像。这个获取大量图像数据并用于特征提取的过程存在大量的细节问题。为爬取能够用于特征抽取的图像，在训练图像获取的过程中需要做如下处理：

[0167] (1)在将爬取的商品图像用于分类训练前，会有一个将不相关图像过滤的步骤，这个步骤往往会过滤大于50%的图像，因此在爬取原始商品图像时，需要按照预计训练图像的两倍以上的规模爬取；

[0168] (2)在从网购平台爬取商品图像时，按照平台所提供统一的规格图像进行爬取，例如分辨率的大致统一和图像格式的统一，在由于图像不一致导致图片分辨率无法统一的情况下，通常平台会保证其最长边一致；

[0169] (3)在应用SURF算法提取特征时，尺寸过小的图像和长宽比例极不协调的图像将会无法提取，因此对于商家提供的这两类图像也需要在爬取过程中避免；

[0170] (4)如上一节所述，所有类别需要保证在类别树中的深度一致。

[0171] 另一方面，在商品图像类别预测的实际应用当中，只提供商品图像的五个可能类别的用户体验相对不佳。因此，本发明在将商品图像的可能类别提供给用户的基础上，自动

从网购平台在线获取相应类别内的相似商品，供用户直接浏览。这种相似性由图像特征提取中所定义。

[0172] 本发明基于从网购平台上获取的真实数据，通过大规模数据的训练，能够自动分析图像中商品的类别信息，向用户提供购物指引，从而简化用户在线购物流程，增强用户体验，在图像检索领域具有广泛的应用价值。

附图说明

[0173] 图1为商品图像类别预测框架流程图。

[0174] 图2为图像四种分辨率的网格划分。

[0175] 图3为基于类别分裂合并的不相关图像过滤算法流程。

[0176] 图4为树结构类层次图(左)与DAG结构类别层次图(右)。

[0177] 图5为商品图像类别预测应用场景图(1)。

[0178] 图6为商品图像类别预测应用场景图(2)。

[0179] 图7为商品图像类别预测应用场景图(3)。

具体实施方式

[0180] 在具体应用中，用户可以点击上传图像按钮，将需要类别预测的图像上传至服务器。这时，服务器将分析图像的基本信息，将图像尺寸、缩略图等信息向用户返回。当用户点击“预测一下”按钮时，系统会自动分析用户所提交的图像内容并预测其类别。当预测完成后，系统向用户返回该商品图像五个可能的类别，并向用户提供8个相关类别的相似商品，供用户选择。

[0181] 当上传一幅蓝色运动鞋的图像时，系统返回板鞋、帆布鞋、运动鞋、休闲鞋和旅游鞋的类别预测，并展示八幅代表对应蓝白相间运动鞋的商品图像。如附图5所示。

[0182] 当上传一幅白色自行车的图像时，系统返回山地自行车、普通自行车、公路自行车、旅行自行车和自行车装备的类别预测，并展示八幅代表对应蓝白相间自行车的商品图像。如附图6所示。

[0183] 当上传一幅粉红色上衣的图像时，系统返回雪纺衫、针织衫、连衣裙、宽松T和针织开衫的类别预测，并展示八幅代表对应粉红色上衣的商品图像。如附图7所示。

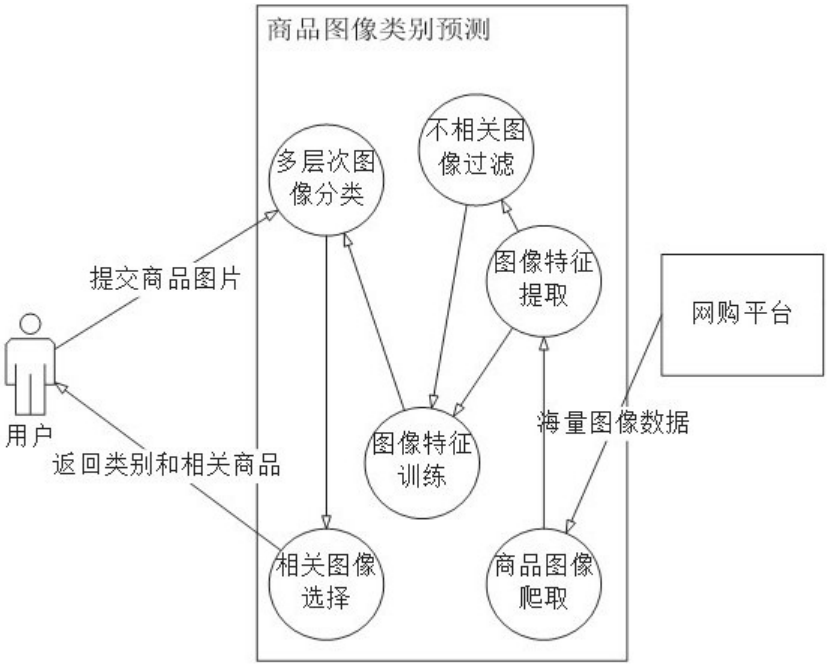

商品图像类别预测
多层次图像分类
不相关图像过滤
图像特征提取
网购平台
提交商品图片
用户
返回类别和相关商品
图像特征训练
海量图像数据
相关图像选择
商品图像爬取

图1

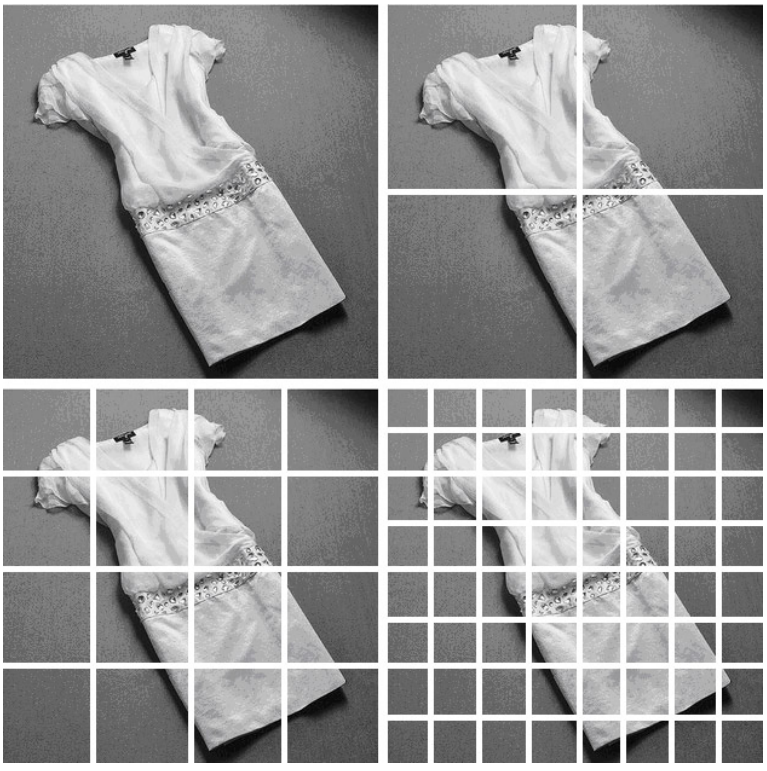


图2


类别分裂合并算法
不相关图像过滤

图3

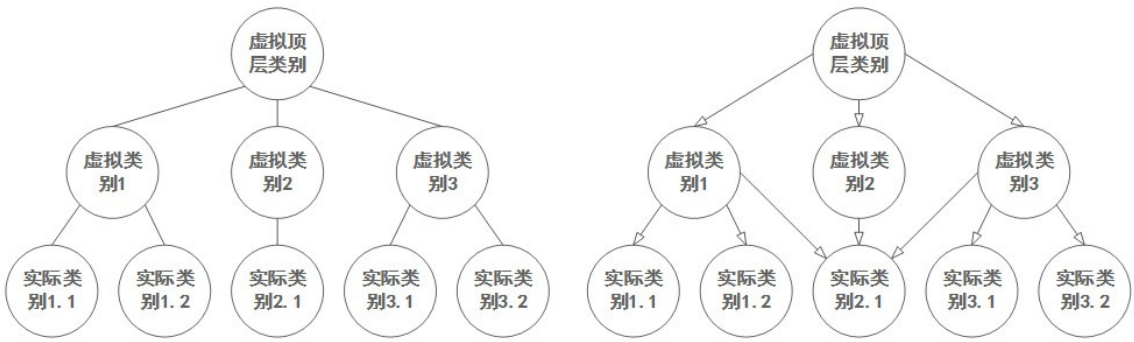

虚拟顶层类别
虚拟类别1
虚拟类别2
虚拟类别3
实际类别1.1
实际类别1.2
实际类别2.1
实际类别3.1
实际类别3.2
虚拟顶层类别
虚拟类别1
虚拟类别2
虚拟类别3
实际类别1.1
实际类别1.2
实际类别2.1
实际类别3.1
实际类别3.2

图4


粘贴图片网址 | 从本地上传
预测一下
您上传的图像尺寸为327×157
该商品图片可能的分类为：
板鞋 帆布鞋 运动鞋 休闲鞋 旅游鞋
您可能感兴趣的商品：

图5


粘贴图片网址 | 从本地上传
预测一下
您上传的图像尺寸为209×137
该商品图片可能的分类为：
山地自行车 普通自行车 公路自行车 旅行自行车 自行车装备
您可能感兴趣的商品：

图6


粘贴图片网址
从本地上传
预测一下
您上传的图像尺寸为 251×267
该商品图片可能的分类为：
雪纺衫 针织衫 连衣裙 宽松T 针织开衫
您可能感兴趣的商品：

图7