**面向多方面密集检索的多粒度感知方面学习模型**

# 孙晓杰

中国科学院大学网络数据科学与技术重点实验室

中国北京sunxiaojie21s@ict.ac.cn

# 杨思惠

中国科学院大学网络数据科学与技术重点实验室

# 毕片

中国科学院大学网络数据科学与技术重点实验室

中国北京bikeping@ict.ac.cn

# 启深张忠义刘冠南张蚂蚁集团

中国北京

# 郭嘉峰\*

中国科学院大学网络数据科学与技术重点实验室

中国北京guojiafeng@ict.ac.cn

# 程雪琪

中国科学院大学网络数据科学与技术重点实验室

中国北京yangsihui22s@ict.ac.cn

## 摘要

{qishen.zqs、zhongyi.lzy、zgn138592}@阿里巴巴-

inc.com

## CCS概念

中国北京cxq@ict.ac.cn

密集检索方法主要集中在非结构化文本上，而较少关注具有各种方面的结构化数据，例如具有诸如类别和品牌等方面的产品。最近的工作提出了两种方法，通过预测与项目方面相关联的值，将方面信息合并到项目表示中，以进行有效的检索。尽管它们很有效，但它们将值视为孤立的类（例如，“智能家居”、“家居、花园和工具”以及“美容和健康”），并忽略了它们的细粒度语义关系。此外，它们要么将方面的学习强制到CLS令牌中，这可能会使其与其表示整个内容语义的指定用途相混淆，要么仅学习具有值预测目标的额外方面嵌入，这可能是不够的，尤其是当项目方面没有注释值时。

arXiv：2312.02538 v2[cs.IR]2024年1月16日

意识到这些限制，我们提出了一个用于多方面密集检索的多粒度感知方面学习模型（MURAL）。它利用跨各种粒度的方面信息来捕获值之间的粗粒度和细粒度语义关系。此外，MURAL结合了单独的方面嵌入作为transformer编码器的输入，使得即使没有方面值注释，屏蔽语言模型目标也可以帮助隐式方面学习。在产品和小程序的两个真实世界数据集上进行的广泛实验表明，MU-RAL的性能明显优于最先进的基线。代码将在URL1提供。

\*郭嘉峰为通讯作者。

1 https://github.com/sunxiaojie99/MURAL



本作品采用知识共享署名国际4.0许可证进行许可。

*WSDM'24，2024年3月4日至8日，墨西哥梅里达*

©2024版权归所有者/作者所有。ACM ISBN 979-8-4007-0371-3/24/03。

https://doi.org/10.1145/3616855.3635770

* 信息系统→信息检索。

## 关键词

密集检索，多方面，预训练

**ACM参考格式：**

孙晓杰、毕克平、郭家峰、杨思慧、张启申、刘忠义、张冠南和程。2024.面向多方面密集检索的多粒度感知方面学习模型。第17届ACM网络搜索和数据挖掘国际会议（WSDM’24）论文集，2024年3月4-8日，墨西哥梅里达。美国纽约州纽约市ACM，9页。https://doi.org/10.1145/3616855.3635770

## 导言

|  |  |  |  |
| --- | --- | --- | --- |
| 查询：“运动手套” | | | **相关性** |
| **项目** | **标题** | **方面：类别（短语级）** |
| **i1** | ATERCEL举重手套全手掌保护、骑自行车、锻炼…… | 运动和健身； |  |
| **i2** | Hestra陆军皮革直升机滑雪手套-经典的3指滑雪雪手套，… | 运动专用服装； |  |
| **i3** | HEAD皮革壁球手套-Web Extra Grip透气手套… | 网球和球拍运动； |  |
| **i4** | HSL 2对可重复使用的厨房洗碗  手套，防水，防滑，园艺，… | 家庭用品； |  |

#### 图1：查询及其候选项的示例。

近年来，密集检索方法在信息检索（IR）和自然语言处理（NLP）社区都得到了广泛的研究[9]。在预训练语言模型（PLM）的肩上，它们已经取得了引人注目的性能。然而，它们大多是针对非结构化数据进行研究的，没有研究如何有效地利用结构化数据的方面信息，如产品的类别和人员的从属关系。例如，在图1中，查询“sports gloves”以运动用手套为目标，因此应避免使用厨房手套。显然，这四个项目的分类可能会有所帮助

区分各种类型的手套，提高检索性能。不幸的是，在密集检索中有效地利用这种方面信息在很大程度上仍未被探索。

最近，Kong等人。[13]提出了两种有效的多方面密集检索模型，即MTBERT和MADRAL。这些方法遵循学习方面嵌入的典型范例，辅助目标是预测它们的相关值【2,3】。一个具体的例子是，图1中i4的方面“cate-gory”的嵌入将通过预测其值，即“家庭用品”来学习。虽然有效，但它们将一个方面的值视为孤立的类，忽略了各种值之间的潜在相关性，这可能导致次优性能。以图1中的项目为例，尽管它们分为四个独立的类别，但前三个与用户查询“运动手套”相关，而最后一个与用户查询“运动手套”无关。预测它们的分类ID的辅助目标平等地对待每个类别，并且可能无法捕获它们的细粒度关系。

注意到这个问题，除了前面考虑的短语级粒度之外，我们建议在更细的粒度上利用方面信息，例如单词和标记级。然后，对于图1中的项目，当我们将它们的类别短语分成小块时，前三者之间的关系将更加清晰，因为它们都有与运动相关的描述，如锻炼、运动、网球等。此外，从语言学的角度来看，句子和短语等较粗的粒度传达更具体的信息，而较细的单元通常携带更一般的信息[22]。由于不同的粒度可以表达不同层次的意图，我们结合了多粒度的方面注释预测来帮助查询/项目表示学习。我们的模型被命名为MURAL，是多粒度感知方面学习模型的缩写。它结合了单独的方面em-

内容令牌之前和CLS之后的床上用品作为

变压器层（如图3所示）。然后，在顶层，用不同粒度级别（例如，短语、单词和令牌）的值预测来监督方面嵌入。与最先进的方法（即MTBERT和MADRAL）相比，MURAL有几个优势（见图2）：首先，与MTBERT混合了CLS中来自项目方面和整体内容语义的信息相比，MURAL分别表示两种类型的信息，并允许它们之间通过门控机制进行更多的交互。第二，与在预训练期间仅学习具有值预测目标的方面嵌入的MADRAL相反，MURAL还指导方面嵌入从掩蔽语言模型损失中学习。即使当项目方面没有注释值时，这也可以帮助隐式方面学习。最后也是最重要的一点，通过跨不同粒度整合方面信息，MURAL可以在不同级别上捕捉方面值之间的语义关系，对检索性能有更大的贡献。

我们在两个具有丰富方面信息的真实世界搜索数据集上进行了广泛的实验。实验结果表明，我们的方法在两个数据集上都显著优于竞争基线。值得注意的是，即使没有方面注释的监督，我们的模型也实现了令人信服的性能，这意味着即使不使用方面信息，也可以通过壁画学习有用的隐式表示。对不同粒度的烧蚀研究表明，每种粒度可以

有助于多方面的检索性能，并将它们结合起来会产生更好的结果。

## 相关工作

密集检索。密集检索模型通常使用双编码器结构进行独立的查询和项目编码，相关性通过简单的相似性函数（如点积）来测量。Karpukhin等[10]用BERT初始化编码器，并将其与批内负片相结合，实现了比早期模型更好的性能。之后，研究人员开始探索各种微调技术来训练更好的密集检索器，包括硬负挖掘[23,31]、知识蒸馏[28]和多向量表示[11,19,33]。例如，熊等人。

[31]提出通过周期性刷新索引的方式在训练过程中动态挖掘硬负。栾等人。[19]通过使用前k个文档令牌嵌入作为项目表示，从不同的角度捕获项目的信息。基于此，Zhang等[33]在项目输入前添加k个特殊标记，得到多向量表示。这些多向量方法旨在从项目中提取多个底层语义信息。相比之下，我们的方法明确地考虑了显式的多方面信息建模。此外，我们的方法只为每个项目输出一个表示向量，节省了索引项目的空间和时间。

最近，Kong等人。[13]介绍了两种将显式方面信息合并到单个表示向量中的方法。第一种方法采用CLS嵌入来同时执行多个方面的方面分类任务。第二种方法将注意力网络添加到PLM中，使其能够分别建模多个方面，然后进行方面融合。它们与我们的方法的差异将在第4节中介绍。

多字段检索。文档中多字段信息（如标题、关键字、描述）的有效利用已经被研究了很长时间。在PLM出现之前，许多神经排序模型被提出来有效地利用项目结构[4,17,32]。例如，Zamani等人。[32]聚合字段级表示以获得项目表示，并采用匹配网络进行最终相关性分数预测。在PLM时代，研究一直集中在多领域信息的利用上[26,27]。例如，Shan等人。[26]提出了字段级局部匹配损失，基于查询和每个文档字段表示进行计算。孙等人。[27]将方面视为文本，并提出了一种有效的预训练方法来捕捉方面和内容文本之间的双向交互。多方面和多字段的区别在于字段包含无限的文本值空间，通常由可变长度的非结构化文本组成。相反，方面具有一组定义的有限值，充当结构化项目的“标签”。鉴于此，它们面临着不同的核心挑战，有效地利用多个方面的信息是一个有价值的研究方向。

预训练的双编码器。研究人员探索了预训练

双编码器架构的检索模型【6，7，14，18，20，30】。例如，Gao和Callan【6】在变压器顶部添加了额外的磁头层，在早期输出和磁头之间具有快捷连接，增强了编码器的CLS嵌入。卢等人。[18]用弱解码器预训练自动编码器

#### 表1：本文中使用的主要符号总结。

**符号 意义**

=(1,2,...,) 查询/项的输入标记序列。

h 输入X的最终表示。

={},=1,..., 一组方面，例如{品牌、颜色、类别}

语言粒度的集合，例如，

方面值嵌入表



品牌 颜色 类别

={},=1,...,

{ℎ,,}

as-的方面值词汇

或，

粒度级别的pect，例如，

=1,...,,=1,...,

={ℎ,ℎ,...}

h或h，



CLS

面向学习的值注释预测

编码器

CLS ！ #…。”



mb！！ ！" ！#

注意注意注意

CLS

！ …。….

#

"

编码器

CLS ！ #…。”

查询或的方面嵌入

粒度上项目的方面。

(a)MTBERT (b)马德拉尔

=1,...,,=1,...,

或，

粒度上的aspect值嵌入表。

#### 图2：SOTA多方面密集检索模型。

像文字或记号一样的颗粒。因此，仅仅依靠短语-

=1,...,,=1,...,

像以前的方法一样，水平值预测可能不会产生

的方面值注释集

一个或一个，

=1,...,,=1,...,

粒度上查询或项的方面

,至⊂

有效的方面表示。例如，如果产品类别值为“手工制品”，则其词级粒度值

用于文档表示学习。与这些针对非结构化数据的预训练方法不同，我们研究了如何在预训练期间将显式方面信息注入编码器表示。在未来，我们将探索如何将我们的方法与现有的研究相结合。

## 预赛

将是“，”，其令牌级粒度值将是“，##，”。形式上，我们将粒度集表示为，其中每个（具有）表示-

发送特定的粒度。在本文中，我们使用三种粒度：=，，。我们用来表示值

∈

[ ]

[ ]

{ }

通过在粒度上分解方面的值获得的词汇。对应的方面值嵌入表变为

∈R×。我们在表1中列出了常用的符号。

序列=1，...，作为输入，并生成上下文

( )

双重编码。标准PLM，例如BERT[5]，采用令牌

表示为：

h(0),h(1)...h()=Φ(),h()∈R, (1)

其中表示隐藏大小，0=是添加到开头的特殊标记。表示H 0通常用作输入的最终表示。在密集检索中，双-

( )

[ ]

编码器架构被广泛采用，其中查询和项

使用PLM分别编码，以获得它们各自的表示向量【16】。然后，使用一个简单的评分函数来计算这两个向量之间的相似度。

方面学习。在密集检索中，方面学习涉及到在以下情况下利用方面信息来提高检索性能

查询或项目与不同方面（例如，产品搜索中的品牌、颜色、类别）相关联。除了内容文本

=1，...，（例如，查询、项目标题），一个查询或项目可以与多个方面相关联，我们将这些方面的集合表示为

( )

={}。为了简单起见，当上下文清楚时，我们省略

## 方法论

在本节中，我们提出了一个用于多方面密集检索的多粒度感知方面学习模型（MURAL），并介绍了它的核心组件。和[13]一样，壁画也是基于伯特[5]。由于MURAL以相同的方式对项目和查询进行编码，因此我们仅使用项目进行说明。

### 方面表示

在预训练模型中合理地表示方面是至关重要的，这样方面学习才能有效地指导它们的训练。为了充分利用Transformer编码器的功能，如图3所示，我们在CLS之后和内容令牌之前引入了几个令牌，以表示不同视角的方面。这与获得CLS的方式一致，并且这些令牌可以与内容令牌充分交互。在预训练期间，当预测内容中的屏蔽令牌时，这些插入的嵌入可以充当上下文的不同视图。这样，这些嵌入也可以从屏蔽语言模型ob-

下标=1。对于每个方面，都存在有限的词汇

在

并从各种隐式中捕获内容语义

方面值，表示为，以及包含每个方面值的嵌入的相应嵌入表R ×。图2显示了两种状态下的方面学习-

∈

艺术多方面密集寻回犬[13]。两种方法都利用内容文本作为编码器输入。具体来说，MTBERT重用CLS来表示方面，而MADRAL通过关注内容的最终层来构建方面的嵌入。这两种方法都通过预测每个方面的相应值注释ID来训练方面嵌入。多粒度。不同粒度的文本字符串捕获

| |

| |

不同层次的语义信息。句子或短语等粗颗粒通常比细颗粒表达更具体的意图

视图，这可以带来更多的好处，尤其是当一个方面没有值注释时。

与以往方法的比较。如图2所示，MTBERT【13】重用CLS令牌来预测项目as-pects的值，这强制CLS将来自项目方面的信息与它最初指定为cap的整体内容语义混合在一起。

真的。两者之间的平衡不能自动学习，CLS可能会对它应该学习什么感到困惑。马德拉尔

[13]通过关注内容令牌的最终表示来分别表示每个项目方面，并通过预测它们的相关值来学习方面嵌入。在预训练期间，对方面嵌入的唯一指导是这个值预测



（

方面值嵌入表

Σ

门控

方面学习

CLS

“！

!!

“”

!!

“！

!"

“”

!"

九月 % &…。#$% #

'

'

$

变压器编码器

“”

!"

CLS

“！

!!

“”

!!

“！

!"

“”

!"

九月

%

&…。#$%

$

#

…

A={！：品牌,"：类别}

G={！：短语,"：单词}

$

输入：更新的苹果MacBook Air 11.6…

$

#### 图3：我们的壁画在两个方面和两个粒度的简单场景中具有基于单目标的分组。

目标，这可能不足以学好它们。更糟糕的是，当没有方面值注释时，它们将不会被更新。相反，在壁画中，方面信息不会与CLS中的整体内容语义混合。通过门控机制，它们可以更多地交互，并且可以自动学习方面的重要性。此外，方面嵌入也可以由掩蔽语言模型损失来指导，这不仅有利于它们的表示学习，而且有利于没有方面值注释的隐式方面学习。

### 方面学习

粒度处的值（与“共享”选项中的初始嵌入相同）。这给了新的嵌入在语义空间中一个体面的起点，并有更好地进行方面学习的自由，其好处将在第6.2节中展示。方面学习目标。一旦我们获得了表示

在粒度上，表示为hR，我们采用

()∈

广泛使用分组对比损失来预训练编码器。它旨在使源表示更接近其目标群体中的实例，同时使其远离其他群体的表示[12]。

1 Σ<UNK> ((h(),e+))

为了简单起见，我们使用粒度为方面的示例

说明。对于壁画的方面学习，有两个重要的方面

L()=−a

, (2)

((h(),e))

+∈A

∈

组成部分：价值表征和方面学习目标。

其中e/e+∈R是嵌入自（·）的纵横值

值表示。通过以下方式进行有效的方面学习

是点积函数，和

根据粗粒度和细粒度预测方面的值注释，值表示起着重要的作用。有两种选择：1）共享主干PLM中现有的令牌嵌入，并通过投影函数计算单词级和短语级颗粒的值嵌入。它可以重用PLM令牌中携带的语义信息。然而，令牌嵌入是朝着PLM和方面学习的目标学习的，这可能会相互干扰。2）声明单独的值嵌入表，这与[13]以及[3]中PLMs时代之前的研究一致。额外的价值嵌入可以服务于模型，在没有其他干预的情况下更好地进行方面学习。然而，如果从头开始训练，这些新参数可能很难优化。就是否与现有编码器共享底层令牌嵌入而言，我们将这两个选项分别称为“共享”和“非共享”。我们在实验中研究了这两种方式（见第6.2节）。

具体来说，对于“共享”选项，我们首先标记每个方面

是粒度上方面的一组方面值注释。

### 多粒度方面分组

假设有方面和粒度，我们的目标是促进每个粒度的每个方面的方面学习，完全是A\*G学习目标。一种直接的方法是使用单个表示来处理这些多个目标。然而，这种方法强制将所有信息一起压缩，严重限制了每个目标的学习能力。因此，我们引入三种分组方案来集成多粒度和多方面：基于单目标的分组、基于粒度的分组和基于方面的分组。

| | | |

基于单一目标的分组。如图3所示，当

只有几个方面和粒度，我们可以直接在输入序列中引入标记来从视图中捕获项目语义。它们中的每一个都说明了多粒度方面组合中的单个目标。指定-

∗

∗

使用BERT标记器的值。然后，我们提取它们的 实际上，我们得到一个序列=(,,..., ，，..，）。

0 1

\*1

来自BERT嵌入表的嵌入，并在令牌嵌入上使用投影仪函数在训练期间获得相应的值嵌入。在本文中，我们采用平均池作为投影函数，因为它简单并且产生

() ∗

我们利用隐藏向量h（=1，...，）作为粒度上从方面的角度表示的项目。方面学习损失函数变为：

在我们的初步实验中，结果与其他方法相似。

1 Σ ţΣ

(3)

对于“非共享”选项，每个方面都有一个单独的嵌入-

每个粒度的ding表，存储其嵌入

()=\*

=1=1

L（）。

。我们不是从头开始训练这些表，而是初始化

价值观

当

∗

很大，添加大量令牌可以

使用PLM中每个表的平均令牌嵌入 对原始输入的语义表示产生不利影响。



品牌短语

品牌词

类别短语 类别词

拉

推

因此，有必要在各种粒度和方面进一步对目标进行分组。

基于粒度的分组。相同的粒度表示相同的语义信息水平，以相同的粒度对目标进行分组是一个合理的选择。在这种情况下，标记将被插入到输入序列中，产生=

| |

0，1，..，，1，2，.，.。它们的编码表示变成h（=1，...，），从粒度上所有方面信息的角度表示项目。单个方面嵌入-

( ) | |

( | )|

ding说明了所有方面的损失L()(=1,...,)

#### 表2：方面相关数据集统计。它显示了预训练语料库中具有非空方面值的查询/项的百分比，以及各种粒度的方面值词汇表大小：短语、单词和标记。

*MA-亚马逊 支付宝*

方面项目 项目/查询品牌 94%（5k、6k、5k） 0.6%/44%（9k、11k、3k）颜色 67%（2k，1k，1k） –

类别 87%（8k、5k、5k） 90%/91%(457,650,548)

粒度的

IR的表示。因此，与[13]类似，我们也采用传销

，即h(=1,...,)对于粒度是相同的

Σ<UNK>

。方面学习目标是：

Σ<UNK>

Σ<UNK>1

作为方面学习之外的预训练目标之一。

1

()=

=1

L()，其中L()=

=1

L（）。 (4)

L()=−

∈()

(\()), (8)

基于方面的分组。另一种选择是按方面对跨多粒度方面的对象进行分组，这样不同的方面信息就不会混合在一起，并且不同级别的粒度可以相互受益。这里，我们在内容令牌之前引入引导令牌：=0，1，..，，1，2，..，.隐藏向量h（=1，...，）捕获对应于方面的输入项的所有粒度的表示。

( ) | |

( | )|

| |

特别地，当使用等式2计算损耗L（）时，

其中表示输入句子，和分别表示屏蔽令牌和来自的剩余令牌。

然后，我们使用方面学习损失与MLM损失联合来预训练Transformer编码器，如下所示，

( ) \ ( )

L()=L()+LA(), (9)

其中是超参数。

方面的表示h

在不同范围内保持一致-

微调。我们采用以下批量softmax交叉en-

tropy损失L作为微调期间的学习目标。注

不同的粒度。在这种聚合方式下，损失 可重新表述如下：

洛杉矶

虽然方面学习损失也可以在微调过程中增加，但我们的实验结果显示没有显著的影响-

1

Σ<UNK>

()=

=1

L()，其中L()=

Σ<UNK>1

=1

L（）。(5)

所有多方面寻回犬的证明。因此，我们在本文中省略了这一目标。

按粒度或方面分组减少了

L

=− ((h,h+))

. (10)

引导代币，容纳场景多面

((h,h+))+−((h,h−))

和粒度。它们的模型架构与图3保持相同，除了相同粒度或方面的方面学习目标是在共享令牌上进行的。

### 方面嵌入融合

出于效率考虑，有必要将多个嵌入式合并为一个嵌入式，以最小化存储和计算成本。受[13]的启发，我们在壁画中采用了“CLS-Gating”融合机制。为了说明融合过程，我们使用第4.3节中讨论的基于单目标的分组方法给出了一个示例。具体来说，我们通过lin-

ear层和softmax函数来计算h（1），...，h（）的加权分数，其中=\*：

w=(h(0)+)∈R, （6）其中R ×和R是可训练参数。然后，我们利用学习到的权重来融合多个嵌入，从而获得输入的最终编码表示：

∈ ∈

Σ<UNK> (7)

h= ·h()。

=1

### 培训目标

预培训。正如在以前的工作[21]中所讨论的，屏蔽语言模型(MLM)[29]任务可以帮助构建好的文本

## 实验设置

### 数据集

我们使用以下两个来自真实世界平台的具有丰富方面信息的大规模搜索数据集进行实验。两个数据集的方面相关统计如表2所示。

#### （1）多方面亚马逊ESCI数据集（MA-Amazon）。马-

亚马逊[27]用项目类别信息丰富了亚马逊ESCI[24]数据集的英文部分。在MA-Amazon中，只有商品有品牌、颜色和类别的注释。项目语料库包含482K个不同的产品，用于预训练。检索数据集分别有17k、3.5 K和8.9 K查询用于微调、验证和测试，没有任何查询重叠。对于每个查询，检索数据集平均提供20.1个项目，以及它们的ESCI相关性判断（精确、替代、补充、不相关），表明每个项目与给定查询的相关性。根据【24】，对于微调和需要二进制标签的度量，我们将Exact视为正值，将所有其他视为负值。（2）支付宝搜索数据集。支付宝是一个中国小程序（类app服务）搜索数据集。在支付宝中，无论是查询还是项目，都标注了品牌和品类两个方面。我们对包含130万个唯一查询的查询语料库和包含180万个不同项目的项目语料库进行预训练。检索数据集分别包含60k、3.3 K和3.3 K个真实用户查询，用于微调、验证和测试，没有查询重叠。请注意

#### 表3：壁画和基线之间的比较。最佳结果（不包括壁画）以粗体显示。、、和分别指示第一/第二组和主干BIBERT中相对于最佳基线的显著改善。

†‡ ∗

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | R@100 | R@500 | NDCG@50 |  | R@100 | R@500 | NDCG@50 |  |
| 比伯特 | 0.6075 | 0.7795 | 0.3929 |  | 0.4464 | 0.6284 | 0.2033 |  |

方法 MA-亚马逊 支付宝

冷凝器 0.6091 0.7801 0.3960 0.4520 0.6423 0.2072

比伯特-康卡特 0.6137 0.7814 0.4005 0.4517 0.6291 0.2103

MTBERT 0.6137 0.7852 0.3969 0.4498 0.6280 0.2064

马德拉尔 0.6088 0.7815 0.3950 0.4506 0.6383 0.2057

墙 0.6282†‡∗ 0.7943†‡∗ 0.4151†‡∗ 0.4556∗ 0.6458‡∗ 0.2046

壁画 0.6371†‡∗ 0.8023†‡∗ 0.4228†‡∗ 0.4630†‡∗ 0.6519†‡∗ 0.2177†‡∗

壁画-CONCAT 0.6389†‡∗ 0.8005†‡∗ 0.4281†‡∗ 0.4669†‡∗ 0.6474‡∗ 0.2124†∗

用于验证和测试的查询不会出现在预训练查询语料库中。相关性数据集中的每个实例都是

<查询，项目，标签>三元组，其中标签指示此查询-项目对的手动注释的二进制相关性。

### 基线

我们采用以下密集检索基线进行比较，包括使用或不使用方面信息的模型：（1）BIBERT【15，25】：标准的双编码器基线和壁画的主干，使用基于BERT的编码器的CLS编码进行查询和项目表示。BIBERT用MLM损失进行预训练，并用损失进行微调（等式10）。（2）Con-Denser[6]：一种为非结构化文本密集检索量身定制的预训练方法。它在预训练期间引入了中间层令牌（不包括CLS）与其对应的头层令牌之间的短路，优化了CLS嵌入以封装更多信息。（3）BIBERT-CONCAT：它将方面值视为文本，并在MLM预训练期间将它们与查询/项目内容连接起来。在微调期间，由于与查询的连接可能会改变查询语义，我们只连接项目方面以进行相关性匹配。（4）MTBERT[13]：基于BIBERT的多任务（MT）学习模型。除了预训练期间的MLM之外，它还使用CLS执行方面预测任务。（5）MADRAL[13]：它结合了一个方面提取注意网络来提取查询和项目的方面表示。这些嵌入在预训练期间从方面预测任务中学习，并在微调期间融合以产生最终表示。（6）壁画、MUR和壁画-CONCAT：壁画是我们提出的多方面密集检索模型。相反，MUR禁用方面学习。具体地，当等式9中的设置为0时，MURAL回归到MUR。MURAL-CONCAT对模型输入采用与BIBERT-CONCAT相同的方面-内容文本连接策略。请注意，除非模型名包括“-CONCAT”，否则模型输入仅由内容文本组成。

L

| |

| |

### 实施细节

我们自己实现了壁画和所有基线，以确保一致的实现细节和公平的比较。

* + 1. 预培训。对于所有方法，编码器对于查询和项目都是共享的，以促进知识共享。具体来说，我们在由项目语料库或查询和项目语料库的混合组成的语料库上进行预训练（当查询方面注释可用时），以获得用于微调的共享编码器。

多粒度值收集。给定一个方面及其在短语级别的原始方面词汇，我们获得其单词和标记粒度词汇：对于单词粒度，我们通过空格和标点符号（对于英语）或使用解霸工具[1]（对于中文）来分割每个方面值，并消除重复以聚合生成的“单词”。对于标记粒度，我们将通过处理每个方面值获得的标记列表与BERT标记器合并，以创建相应的词汇集。

模型预训练。我们使用Google的公共检查点初始化所有BERT组件，并使用Adam优化器和线性预热技术。MA-Amazon/支付宝数据集的学习率和纪元分别设置为1e-4/5e-5和20/10。最大令牌长度为156，MLM掩码比率对于项目为0.15，对于查询为0.3。对于所有需要调整训练目标缩放系数的方法，我们根据微调后的验证集性能统一选择系数。这些系数以0.1为间隔从0.1到1变化。对于我们的方法，我们在等式9中设置为0.1。我们使用以下微调过程每两个时期评估预训练的模型检查点，并在验证集上选择最佳检查点。

* + 1. 微调。对于这两个数据集，我们使用Tevatron工具包[8]微调了20个时期的所有模型。继之前的工作[10]之后，除了批内否定之外，我们还为每个查询包括一个硬否定样本。我们使用5e-6的学习率和

64.查询的最大标记长度设置为32，项目的最大标记长度为156。所有模型都用相关性损失L（等式10）进行训练。

* + 1. 评估指标。我们报告R@100、R@500和NDCG@50。根据【24】，我们将1.0、0.1、0.01和0.0的增益分别分配给MA-Amazon的E、S、C和I。我们进行双尾配对t检验（p<0.05）以确定显著差异。

## 实验结果

### 整体表现

我们将壁画与利用和不利用方面信息的基线模型进行比较。如表3所示，我们有以下观察结果：（1）利用方面信息的模型（除了BIBERT、Condenser和MUR之外的方法）在不使用它的情况下优于它们的主干（BERT）。在多向密集寻回犬中，壁画表现最好，差距非常大。这证实了在查询/项目表示学习中结合方面的必要性。（2）在MA-Amazon上，MADRAL的表现不如更简单的MTBERT。我们认为这是由于MA-Amazon的预训练数据较少以及缺乏

#### 表4：马亚马逊上壁画的变体。表明与最佳选项的显著差异。

†

方法 R@100 R@500 NDCG@50

壁画ℎ\_ 0.6336†0.8005 0.4195†壁画ℎ\_ 0.6371 0.8023 0.4228壁画ℎ\_ 0.6340†0.8003†0.4195†

壁画ℎ\_ 0.6300†0.7980†0.4171†壁画ℎ\_ 0.6333†0.7996†0.4184†壁画ℎ\_ 0.6321†0.7995†0.4173†

壁画ℎ\_0.6337†0.8006 0.4196†壁画\_ 0.6141†0.7873†0.4009†壁画\_\_ 0.6212†0.7890†0.4064†

查询的方面注释，这使得MADRAL的方面嵌入没有被充分学习。相比之下，壁画始终取得引人注目的表现。（3）Condenser是一种用于非结构化文本检索的更高级的预训练模型，有时会优于基线多方面密集检索器。值得注意的是，高级预训练技术的收益与我们的方法正交。我们的方法可以很容易地集成到更强的主干中，如冷凝器，并可以实现更好的性能。我们把它留在以后的工作中。（4）BIBERT-CONCAT在两个数据集上的一些指标方面比MTBERT和MADRAL表现更好，这表明将方面连接为文本字符串可能是有益的。然而，为了获得良好的性能，在相关性匹配期间应该特别注意查询方面。在同一模型中结合级联和方面预测（MURAL-CONCAT）并不总是比没有级联时产生更好的性能。我们对MTBERT和MADRAL也有类似的观察，但由于空间问题，我们没有报告它们。原因可能是当使用方面作为模型输入和学习目标时，模型学习不想要的快捷方式。（5）即使不使用方面注释（MUR），我们的方法也显示出竞争性能。除了支付宝上的NDCG@50之外，MUR在所有指标上都优于大多数基线模型。这表明MUR可以从隐式视角捕获互补信息以用于最终表示。这也证实了方面表征和MLM训练在壁画方面的优势。

### 模型变体研究

在这一小节中，我们研究壁画中基本成分的各种选择。为了再现性，所有实验都在公共MA-Amazon数据集上进行。

价值表征研究。在表4中，我们观察到使用独立的值嵌入空间（第4.2节中的“非共享”选项）会带来更好的性能。正如我们前面提到的，“共享”选项优化了BERT和方面学习中的目标的令牌嵌入，这可能会相互干扰并限制模型在方面预测上的能力。然而，在“非共享”选项下，可能很难从头开始优化单独的值嵌入，而其他参数只需要微调。为了查看这是否会影响模型性能，我们不使用与“共享”选项相同的初始状态，而是在

#### 表5：MA-Amazon数据集上壁画的类别和粒度消融研究。，表明与壁画和BIBERT相比有显著差异。

方法 R@100 R@500 NDCG@50

†‡

比伯特 0.6075 0.7795 0.3929

壁画 0.6371 0.8023 0.4228

壁画 0.6289†‡0.7951†‡0.4168†‡壁画 0.6284†‡0.7942†‡0.4158†‡壁画0.6315†‡0.7994†‡0.4166†‡

壁画ℎ 0.6315†‡0.7983†‡0.4185†‡壁画 0.6309†‡0.7994†‡0.4194†‡壁画 0.6305†‡0.7982†‡0.4192†‡

保留壁画中的其他最佳设置。表4中壁画\_的受损性能证实了我们的假设，并显示了体面初始化状态的好处。

分组方法研究。在表4中，我们观察到MU-

RAL\_按粒度对目标进行分组，表现最好。请注意，在支付宝数据集上，它有更少的

方面，壁画\_，基于单目标的分组，具有最佳性能。这与我们在第4.3节中的主张一致，即随着方面计数的增加，进一步分组有利于模型训练。基于这些观察，我们建议：

（1）对于少量的方面和粒度，只需对每个目标使用独立学习（壁画\_）。(2)

当有更多的方面和粒度时，将多个目标分组在一个指导令牌中可能是更好的选择。

引导令牌与融合方法研究。我们没有为方面学习添加单独的指导标记，而是研究了一种在输入序列开始时重用相同数量的标记来进行方面学习的变体，表示为壁画

\_。结果表明，壁画\_的性能与表3中的最佳基线相似或更好，但明显差于壁画的最佳变体。这表明多粒度感知方面学习是有益的，但是需要使用单独的引导令牌来进行学习。

为了研究CLS门控（在第4.4节中介绍）是否有助于方面嵌入融合，在壁画\_\_中，我们重新-

移动它并使用CLS嵌入作为最终表示。CLS自然地融合倒数第二层中的方面嵌入，而方面学习在最后一层中进行。该变体比表3中的最佳基线表现更好，但比最佳变体差。这表明融合应该用适当的加权机制对最终的方面嵌入进行。

### 消融研究

我们消除了多方面、多粒度和查询/项目方面学习的各种组件。在本节中，我们的实验也基于丰富的MA-Amazon数据集。此外，我们验证了查询和项目副作用对支付宝数据集的重要性，因为MA-Amazon缺乏查询端方面的信息。

方面和粒度的影响。在表5中，我们首先研究了壁画中多方面和多粒度的影响。我们发现：（1）每个方面都有助于模型性能，尤其是类别，与[13]一致。（2）单独的各粒度

#### 表6：支付宝上查询和项目方面的消融。，表明与壁画和BIBERT相比有显著差异。

†‡

方法 R@100 R@500 NDCG@50

比伯特 0.4464 0.6284 0.2033

壁画 0.46300.6519 0.2177

壁画− 0.4569‡ 0.6400†‡ 0.2126†‡壁画− 0.4573‡ 0.6454†‡ 0.2103†‡

马德拉尔

c1：运动专用服装c2：运动与健身

c3：家庭用品

壁画

c1：运动专用服装c2：运动与健身

c3：家庭用品



#### 表7：支付宝数据集上的类别方面准确度。



方法

马德拉尔

壁画

查询

doc

短语

短语

训练前0.86

0.88

微调0.11

0.81

训练前0.97

0.98

微调0.20

0.87

短语

0.89

0.85

代币

word

0.88

0.89

0.82

0.98

0.73

0.97

0.93

0.97

0.80

0.93

有利于模型性能，将它们结合起来会产生更好的结果。不同的粒度捕获不同级别的语义，它们可以相互补充。

查询/项目方面的影响。我们禁用方面学习

在表6中的查询/项目侧。我们观察到查询和项目方面都是有益的，查询方面有更大的影响，这也与【13】一致。这并不奇怪，因为查询方面是从诸如意图分类的查询分析中获得的，并且携带更多附加信息。

### 方面学习精度

在表7中，我们比较了预训练和微调后壁画与基线方法的准确性，以更好地理解方面学习过程。我们只分析最重要的方面

-支付宝上的分类。MA-亚马逊等方面也有类似的结论。考虑到每个项目可能有多个类别注释，我们使用accuracy@3来计算准确性。查询和项目方面预测的评估分别基于支付宝数据集的测试查询集和项目语料库。

首先，所有方法在预训练中进行方面学习后都具有高精度，而在微调后精度较低。由于我们仅在微调期间使用相关性损失，因此预计精度会下降。在我们的实验中，我们发现在微调期间添加方面学习损失可以提高方面预测的准确性，但会损害检索性能。我们推测，这个目标将模型参数引导到与相关性匹配目标不一致的地方。因此，更高的方面预测精度并不总是与更好的检索性能同时发生。

其次，微调后MTBERT的预测精度急剧下降。由于MTBERT使用相同的CLS令牌来进行相关性匹配和方面预测，因此在微调期间仅针对相关性匹配进行优化会破坏其预测方面值的能力。相比之下，MADRAL和MURAL在微调后保留了大部分这种能力，因为它们使用额外的方面嵌入来执行方面学习。

最后，对于短语级评估，壁画具有最好的方面预测准确性。我们知道，壁画也具有最好的检索性能，这意味着壁画可以很好地学习这两个目标，并让更好的方面嵌入辅助相关性匹配

#### 图4：MA-Amazon上MADRAL和MURAL的项目表示的t-SNE图。

更多。值得注意的是，单词和标记级别的准确性无法与短语级别相比，因为基本事实是不同的。更精细级别的预测精度也不错。当粒度变细时，微调后精度变低，这可能是因为更细的粒度具有更多的地面真值，使多标签分类更具挑战性。

### 案例可视化

我们可视化三个类别的项目表示，如图1所示，以查看它们在语义空间中的分布。具体来说，c1（运动专用服装）和c2（运动和健身）在语义上相似，而c3（家庭用品）与前两者无关。我们首先使用MADRAL和MURAL来获取MA-Amazon上的所有项目表示，并将项目归入它们的类别。然后，我们随机抽取c1、c2和c3的20个项目，并使用图4中的t-SNE工具包绘制它们。我们可以观察到，MADRAL将c1、c2和c3分开的程度相似。相比之下，壁画将相关类别c1和c2放置得更近，而将它们与不相关的c3放置得更远。这表明MADRAL无法辨别c1和c2之间的语义相似性，因为它将不同的短语级产品类别视为孤立的id，忽略了它们的单词级语义联系。壁画可以捕捉相似方面值之间的细粒度语义关系，同时通过结合粗粒度和细粒度信息来保持精确的短语级方面区分。

## 结论

在本文中，我们提出了一种多粒度感知的方面学习模型，该模型增强了结构化数据中附加方面信息的利用。与以前忽略不同方面值之间语义关系的方法不同，我们的方法结合了方面值的多个粒度来促进查询/项表示学习。通过有效地从隐式视图中捕获查询/项的语义，我们的模型即使没有方面注释的监督也能实现令人信服的性能。两个真实世界数据集的实证结果证明了壁画的优越性。

## 致谢

本工作得到了国家自然科学基金（NSFC）62302486号、ICT CAS创新项目E361140号、CAS特别研究助理资助项目、联想-CAS联合实验室青年科学家项目和JCKY2022130C039号项目的资助。这项工作也得到了蚂蚁集团通过Ant创新研究计划的支持。

## 参考文献

1. 2023.https://github.com/fxsjy/jieba。
2. 艾庆耀、瓦希德·阿齐兹、陈旭和张永峰。2018.学习异构知识库嵌入以进行可解释推荐。算法11，9（2018），137。
3. 艾庆耀、张永峰、毕和W.布鲁斯·克罗夫特。2020.具有动态关系嵌入模型的可解释产品搜索。ACM Trans。信息。系统。38,1(2020),4:1–4:29.
4. 赛义德·巴拉内辛科尔丹、亚历山大·科托夫和费多尔·尼古拉耶夫。2018.用于即席结构化文档检索的注意力神经架构。第27届ACM信息与知识管理国际会议论文集，CIKM 2018，意大利都灵，2018年10月22-26日，Alfredo Cuzzocrea、James Allan、Norman W.Paton、Divesh Srivastava、Rakesh Agrawal、Andrei Z.Broder、Mohammed J.Zaki、K.Sel ç uk Candan、Alexandros Labrinidis、Assaf Schuster和Haixun Wang（编辑）。美国计算机学会，1173-1182。
5. 雅各布·德夫林、张明伟、肯顿·李和克里斯蒂娜·图塔诺娃。2018.Bert：用于语言理解的深度双向Transformers的预训练。

CoRR abs/1810.04805(2018).arXiv：1810.04805

1. 高鲁豫和杰米·卡兰。2021.Condenser：用于密集检索的预训练架构。2021年自然语言处理经验方法会议论文集，EMNLP 2021，虚拟事件/多米尼加共和国蓬塔卡纳，2021年11月7-11日，Marie-Francine Moens，Huang，Lucia Specia和Scott Wen-tau Yih（编辑。）.计算语言学协会，981-993。
2. 高鲁豫和杰米·卡兰。2022.密集段落检索的无监督语料库感知语言模型预训练。计算语言学协会第60届年会论文集（第1卷：长论文），ACL 2022，爱尔兰都柏林，2022年5月22-27日，Smaranda Muresan，Preslav Nakov和Aline Villavicencio（编辑。）.计算语言学协会，2843-2853。
3. 高鲁豫、马学光、林志颖和杰米·卡兰。2022.Tevatron：高效灵活的密集检索工具包。更正abs/2203.05765(2022)。arXiv：2203.05765
4. 郭家峰、蔡银琼、范宜兴、孙飞、张如清和程。2022.第一阶段检索的语义模型：综合综述。ACM Trans。信息。系统。40,4(2022),66:1–66:42.
5. Vladimir Karpukhin、Barlas Oguz、Sewon Min、Ledell Wu、Sergey Edunov、Danqi Chen和Wen-tau Yih。2020.开放领域问答的密集段落检索。CoRR abs/2004.04906(2020).arXiv：2004.04906
6. 奥马尔·哈塔布和马泰·扎哈里亚。2020.科尔伯特：通过BERT上的情境化后期交互进行高效和有效的段落搜索。第43届国际ACM SIGIR信息检索研究与发展会议论文集，SIGIR 2020，虚拟活动，中国，2020年7月25-30日，Jimmy X.Huang，Yi Chang，Xueqi Cheng，Jaap Kamps，Vanessa Murdock，Ji-Rong Wen和Yiqun Liu（编辑。）.ACM，39–48。
7. Prannay Khosla、Piotr Teterwak、Chen Wang、Aaron Sarna、Yonglong Tian、Phillip Isola、Aaron Maschinot、Ce Liu和Dilip Krishnan。2020.监督对比学习。《神经信息处理系统进展33：2020年神经信息处理系统年会》，NeurIPS 2020，2020年12月6-12日，虚拟，Hugo Larochelle，Marc’Aurelio Ranzato，Raia Hadsell，Maria-Florina Balcan和Hsuan-Tien Lin（编辑。）.
8. 孔维泽、斯瓦拉吉·卡丹加、李成、莎琳·库马尔·古普塔、张明阳、徐文松和迈克尔·本德斯基。2022.多方位密集检索。在KDD’22：第28届ACM SIGKDD知识发现和数据挖掘会议，美国DC华盛顿州，2022年8月14-18日，Aidong Zhang和Huzefa Rangwala（编辑。）.美国计算机学会，3178-3186。
9. Kenton Lee，Ming-Wei Chang和Kristina Toutanova。2019.弱监督开放域问答的潜在检索。《计算语言学协会第57届会议论文集》，ACL 2019，意大利佛罗伦萨，2019年7月28日至8月2日，第1卷：长篇论文，Anna Korhonen、David R.Traum和Lluís Màrquez（编辑）。计算语言学协会，6086-6096。
10. 林志颖、罗德里戈·诺盖拉和安德鲁·耶茨。2021.用于文本排名的预训练变形金刚：BERT和Beyond。摩根&克莱普尔出版社。
11. 林志颖、罗德里戈·弗拉塞托·诺盖拉和安德鲁·耶茨。2021.用于文本排名的预训练变形金刚：BERT和Beyond。摩根&克莱普尔出版社。
12. 刘斌生、陆晓璐、柳文欢·库尔兰和J·谢恩·卡尔佩珀。2018.利用基于字段的相关性建模提高搜索效率。第23届澳大拉西亚文档计算研讨会论文集，ADCS 2018，新西兰达尼丁，2018年12月11-12日。ACM，11：1-11：4。
13. 陆舒琪、熊晨燕、何迪、柯国林、瓦利德·马利克、窦志成、保罗

贝内特、刘铁燕和阿诺德·奥弗维克。2021.少即是多：使用弱解码器预训练强连体编码器。CoRR abs/2102.09206(2021)。arXiv：2102.09206

1. 易栾、雅各布·爱森斯坦、克里斯蒂娜·图塔诺娃和迈克尔·柯林斯。2021.文本检索的稀疏、密集和注意表示。反式。协会。计算机。语言学9（2021），329–345。
2. 马新宇、郭家峰、张如清、范、程。2022.通过对比跨度预测预训练用于密集检索的判别文本编码器。SIGIR’22：第45届国际ACM SIGIR信息检索研究与开发会议，西班牙马德里，2022年7月11日至15日，Enrique Amigó、Pablo Castells、Julio Gonzalo、Ben Carterette、J.Shane Culpepper和Gabriella Kazai（编辑）。美国计算机学会，848–858。
3. 马新宇、郭家峰、张如清、范、向吉和程。2021.PROP：用于即席检索的代表词预测预训练。第14届ACM网络搜索和数据挖掘国际会议论文集（2021）。
4. 鲁图·穆尔卡-梅塔、杰里·R·霍布斯和爱德华·H·霍维。2011.粒度

自然语言话语。第九届计算语义学国际会议论文集，IWCS 2011，2011年1月12-14日，英国牛津，Johan Bos和Stephen Pulman（编辑。）.计算机语言学协会。

1. 曲英奇、丁、刘静、刘凯、任瑞阳、赵欣、董大喜、吴华和王海峰。2021.RocketQA：开放域问答密集段落检索的优化训练方法。计算语言学协会北美分会2021年会议论文集：人类语言技术，NAACL-HLT 2021，在线，2021年6月6-11日，Kristina Toutanova，Anna Rumshisky，Luke Zettle-Moyer，Dilek Hakkani-Tür，Iz Beltagy，Steven Bethard，Ryan Cotterell，Tanmoy Chakraborty和Yichao Zhou（编辑）。计算语言学协会，5835-5847。
2. Chandan K.Reddy、Lluís Màrquez、Fran Valero、Nikhil Rao、Hugo Zaragoza、Sambaran Bandyopadhyay、Arnab Biswas、Anlu Xing和Karthik Subbian。2022.购物查询数据集：用于改进产品搜索的大规模ESCI基准。更正abs/2206.06588(2022)。arXiv：2206.06588
3. 尼尔斯·赖默斯和伊琳娜·古列维奇。2019.句子-BERT：使用暹罗BERT网络的句子嵌入。2019年自然语言处理经验方法会议和第九届自然语言处理国际联合会议论文集，EMNLP-IJCNLP 2019，中国香港，2019年11月3-7日，Kentaro Inui，Jing Jiang，Vincent Ng和Xiaojun Wan（编辑。）.计算语言学协会，3980-3990。
4. 单宏宇、张启申、刘忠义、张冠南和李。2023.超越双塔：候选检索的属性引导表征学习。ACM网络会议论文集2023.3173-3181。
5. 孙晓杰、毕克平、郭家峰、马新宇、范宜兴、山宏宇、张启申和刘忠义。2023.用于多方面密集检索的方面-内容文本相互预测预训练。第32届ACM信息和知识管理国际会议论文集（英国伯明翰）（CIKM’23）。计算机械协会，纽约，

美国纽约，4300-4304。https://doi.org/10.1145/3583780.3615157

1. Amir Vakili Tahami、Kamyar Ghajar和Azadeh Shakery。2020.为基于快速检索的聊天机器人提取知识。第43届国际ACM SIGIR信息检索研究与发展会议论文集，SIGIR 2020，虚拟活动，中国，2020年7月25-30日，Jimmy X.Huang，Yi Chang，Xueqi Cheng，Jaap Kamps，Vanessa Murdock，Ji-Rong Wen和Yiqun Liu（编辑。）.ACM，2081–2084。
2. 威尔逊·泰勒。1953.《完形填空程序》：测量可读性的新工具。

新闻与大众传播季刊30（1953），415-433。

1. 肖石涛、刘正、邵和曹赵。2022.RetroMae：通过屏蔽自动编码器预训练面向检索的语言模型。2022年自然语言处理经验方法会议论文集，EMNLP 2022，阿拉伯联合酋长国阿布扎比，2022年12月7-11日，Yoav Goldberg、Zornitsa Kozareva和张悦（编辑。）.计算语言学协会，538–548。
2. Lee Xiong、Chenyan Xiong、Ye Li、Kwok-Fung Tang、Jialin Liu、Paul N.Bennett、Junaid Ahmed和Arnold Overwijk。2021.密集文本检索的近似最近邻负对比学习。第九届学习表征国际会议，ICLR 2021，虚拟活动，2021年5月3日至7日。OpenReview.net.
3. 哈米德·扎马尼、巴斯卡尔·米特拉、宋夏、尼克·克拉斯威尔和索拉布·蒂瓦里。2018.具有多个文档字段的神经排序模型。第11届ACM网络搜索和数据挖掘国际会议论文集，WSDM 2018，美国加利福尼亚州玛丽娜德尔雷，2018年2月5-9日，易昌，翟成祥，刘燕和Yoelle Maarek（编辑。）.美国计算机学会，700–708。
4. 张舜宇、梁耀波、龚明清、姜大新和段南。2022.面向开放领域密集检索的多视图文档表示学习。计算语言学协会第60届年会论文集（第1卷：长论文），ACL 2022，爱尔兰都柏林，2022年5月22-27日，Smaranda Muresan，Preslav Nakov和Aline Villavicencio（编辑。）.计算语言学协会，5990-6000。