

基于对比学习的大型语言模型反向词典任务提示生成方法

田思成, 黄少滨, 王锐, 李熔盛, 杜治娟

引用本文

田思成, 黄少滨, 王锐, 李熔盛, 杜治娟. 基于对比学习的大型语言模型反向词典任务提示生成方法[J]. 计算机科学, 2024, 51(8): 256-262.

TIAN Sicheng, HUANG Shaobin, WANG Rui, LI Rongsheng, DU Zhijuan. [Contrastive Learning-based Prompt Generation Method for Large-scale Language Model ReverseDictionary Task](#) [J]. Computer Science, 2024, 51(8): 256-262.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于熵值过滤和类质心优化的无监督域适应](#)

Unsupervised Domain Adaptation Based on Entropy Filtering and Class Centroid Optimization

计算机科学, 2024, 51(7): 345-353. <https://doi.org/10.11896/jsjcx.230500144>

[融入多影响力与偏好的图对比学习社交推荐算法](#)

Graph Contrastive Learning Incorporating Multi-influence and Preference for Social Recommendation

计算机科学, 2024, 51(7): 146-155. <https://doi.org/10.11896/jsjcx.230400147>

[基于句信息增强词信息的方面级情感分类](#)

Aspect-based Sentiment Classification for Word Information Enhancement Based on Sentence Information

计算机科学, 2024, 51(6): 299-308. <https://doi.org/10.11896/jsjcx.230600059>

[基于对比学习的视觉增强多模态命名实体识别](#)

Vision-enhanced Multimodal Named Entity Recognition Based on Contrastive Learning

计算机科学, 2024, 51(6): 198-205. <https://doi.org/10.11896/jsjcx.230400052>

[结合图对比学习的多图神经网络会话推荐方法](#)

Graph Contrast Learning Based Multi-graph Neural Network for Session-based RecommendationMethod

计算机科学, 2024, 51(5): 54-61. <https://doi.org/10.11896/jsjcx.230300092>

基于对比学习的大型语言模型反向词典任务提示生成方法

田思成¹ 黄少滨¹ 王锐¹ 李熔盛¹ 杜治娟^{2,3}

¹ 哈尔滨工程大学计算机科学与技术学院 哈尔滨 150001

² 生态大数据教育部工程研究中心 内蒙古 010021

³ 内蒙古大学计算机学院 内蒙古 010021

(standby@hrbeu.edu.cn)

摘要 反向词典任务是一种新兴的任务,目的是根据给定的定义来查找对应的单词。大规模语言模型为这一任务提供了新的可能性,但是提示语句的质量会影响大模型的性能。为此,提出了一种基于对比学习的提示生成方法。该方法在从多个语义层面上理解定义语义的同时,还利用对比学习的原理在训练过程中引入了负例,提升了模型的泛化能力。通过这种方法,可以将目标单词缩小到一个小范围内,然后用大模型从这个范围内选择最符合定义语义的单词。实验结果表明,该方法可以有效地提升大规模语言模型在反向词典任务上的表现。提示生成模型有 94.7% 的概率生成包含目标词的范围,大规模语言模型有 58.03% 的概率直接选出目标单词,有 74.55% 的概率在给出 5 个候选单词时包含目标单词。

关键词: 反向词典;大规模语言模型;对比学习;多个语义层面;对比损失

中图分类号 TP391

Contrastive Learning-based Prompt Generation Method for Large-scale Language Model Reverse Dictionary Task

TIAN Sicheng¹, HUANG Shaobin¹, WANG Rui¹, LI Rongsheng¹ and DU Zhijuan^{2,3}

¹ College of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China

² Engineering Research Center of Ecological Big Data, Ministry of Education, Inner Mongolia, 010021, China

³ College of Computer, Inner Mongolia University, Inner Mongolia, 010021, China

Abstract Reverse dictionary task is an emerging task that aims to find the corresponding word based on a given definition. Large-scale language models offer new possibilities for this task, but the quality of the prompt sentences affects the performance of the large models. To this end, this paper proposes a contrastive learning-based prompt generation method. This method extracts definition semantics from multiple semantic levels. It also enhances the model's generalization ability by incorporating negative examples through contrastive learning. With this method, we can narrow down the target word to a small range, and use a large model to select the most semantically consistent word from this range. Experimental results show that the proposed method can effectively improve the performance of large-scale language models on the reverse dictionary task. The prompt generation model has a 94.7% probability of generating a range that contains the target word. The large-scale language model has a 58.03% probability of directly selecting the target word, and a 74.55% probability of including the target word when five candidate words are given.

Keywords Reverse dictionary, Large-scale language model, Contrastive learning, Multiple semantic scales, Contrastive loss

1 引言

反向词典是一种通过输入定义来检索匹配单词的任务,它与传统的词典正好相反。这个任务最早是在 SemEval-2022 Task 1 “CODWOE: Comparing Dictionaries and Word Embeddings”^[1]中提出的。Qi 等^[2]开发了一个在线的反向词典系统,用户可以输入对单词含义的描述,然后得到一个满足

该描述的单词列表。反向词典可以帮助人们在遗忘某个单词或者想要寻找与某个概念相关的单词时,快速地找到所需的结果。反向词典不仅可以用来应对短暂时遗忘、缓解“网络失语症”、辅助语言学习者掌握词汇,还可以用来进行创意写作、玩填字游戏、回答基础知识问题等。

大规模语言模型为此任务的解决带来了突破,但是提示语句的质量会影响大模型的性能。不恰当的提示语句可能会

到稿日期:2023-06-26 返修日期:2023-11-14

基金项目:生态大数据教育部工程研究中心开放课题

This work was supported by the Open Project of Engineering Research Center of Ecological Big Data, Ministry of Education.

通信作者:黄少滨(huangshaobin@hrbeu.edu.cn)

导致错误的结果。如图 1 所示,本文对大规模语言模型 GPT3 在不同提示语句下的表现进行了测试。单词‘concession(让步)’的定义是‘action of conceding or granting something(做出让步或给予某物的行为)’。如果用‘I want to know ‘the action of conceding or granting something points to which word?’作为提示语句,得到的结果是‘The word that points to the action of conceding or granting something is “concede” or “grant”’。但如果用‘The meaning of that word in the following words is ‘the action of conceding or granting something’, salaam, concession, cryobiology, goof-off, conditional, chippy, now, macho, smite and squeeze?’作为提示语句,得到的结果是‘concession’。

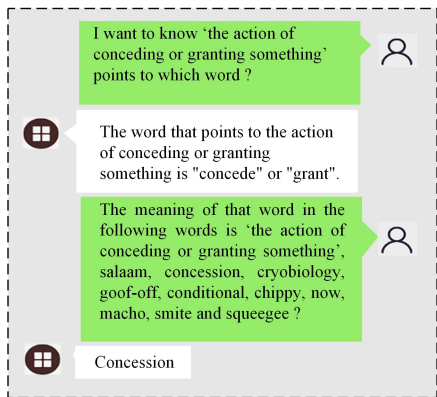


图 1 大模型对话样例

Fig. 1 Example of large model dialogue

可以发现,如果给出一个包含候选词集的提示语句,则可以降低大规模语言模型在反向词典任务上的难度,从而得到更准确的答案。但是,用户要想编写这样的提示语句,需要具备大量的先验知识。如何用人工智能的方法自动实现对提示语句的改进是一个研究难点。本文的研究目标是通过提出一种基于对比学习的提示语句生成方法,自动生成高质量的提示语句。该方法分为两个步骤:第一步是建立一个基于对比学习的反向词典模型(Reverse Dictionary Model Based on Contrastive Learning, RDMCL),输出前 n 个最符合定义语义的单词。第二步是利用这 n 个单词作为候选词集,创建一个高质量的提示语句。

现有的反向词典模型主要考虑单一级别的语义信息^[3-4],未能充分利用各种神经网络在提取不同级别的语义信息上的优势。例如,循环神经网络可以提取上下文特征,能用于提取句子级别的语义信息,但是会面临梯度消失的问题。卷积神经网络可以学习句子中的局部特征,能用于提取短语级别的语义信息,但是缺乏位置信息。注意力机制可以放大句子中关键词的影响,但是无法捕捉序列中的顺序信息。相比之下,RDMCL是一个从多个语义层面上理解定义语义的模型。它利用双向长短期记忆网络(Bidirectional Long Short-Term Memory, Bi-LSTM)捕捉句子级别的语义信息;利用卷积神经网络捕捉短语级别的语义信息;利用自注意力机制捕捉单词级别的语义信息。然后,将这 3 种不同层级的语义信息融合在一起,并将其映射到目标单词的语义空间中。相比部分现有反向词典方法仅对某一类别的语义信息进行提取,RDMCL既提取了句子的顺序信息,又学习了句子的局部特征,并且还放大了关键词的影响。

为了提高反向词典模型的泛化能力,RDMCL提出了一种新颖的有监督对比损失函数。有监督对比学习可以利用标签信息将嵌入空间中的一个锚点和一个正例样本拉近,将锚点与负例样本拉远,从而改进模型在未见过数据上的表现^[5-6]。然而,现有的有监督对比损失函数大多是针对分类任务的,对于反向词典这类回归任务的研究较少。因此,RDMCL构建了一个针对于回归任务的有监督对比损失函数,其目的是让定义的语义表示不仅与目标单词接近,而且与其他无关单词远离。通过 RDMCL 可以给出 n 个最有可能代表定义语义的单词。将带有候选词表的提示语句输入大规模语言模型,得到最终的结果。

本文在文献[7]提出的 Oxford 词典数据集的基础上进行了修改,构建了一个新的数据集。Oxford 数据集由单词、定义和上下文组成,然而文献[8]指出该数据集中存在很多噪声,如上下文中缺少目标单词。为此对数据集进行了清洗,得到了一个 New Reverse Dictionary 数据集。在该数据集上的实验结果表明,当 $n=150$ 时,有近 95% 的概率使得目标单词出现在候选词集中。利用高质量的提示语句可以使得大规模语言模型直接给出正确答案的概率为 58.03%,给出 5 个候选单词时包含目标单词的概率为 74.55%。本文的主要贡献如下:

1)提出了基于对比学习的反向词典模型 RDMCL,从多个语义层面上理解定义语义。

2)提出了一种有监督对比损失函数,将负例引入反向词典任务中,提升了模型性能,在多个指标上超越了基准模型。

3)基于 RDMCL 给出的候选词表构建高质量提示语句,提升了大模型在反向词典任务上的性能。

2 相关工作

本文旨在自动地生成高质量的提示语句。为了实现这一目标,提出了一个基于对比学习的反向词典模型 RDMCL。本章将重点介绍构建 RDMCL 所用的反向词典方法和对比学习方法。

2.1 反向词典

目前,已经有多位研究者针对在 SemEval-2022 任务^{1[1]}中首次提出的反向字典任务进行了研究。文献[3]提出了一个面向西班牙语的语言驱动系统。该系统能够利用解释生成词汇,并从词典学角度分析数据集中存在的一致性和准确性缺陷;从而给出了一种预处理方法,避免了数据集在模型中可能带来的问题。一些研究将单一语言泛化到多种不同语言的研究中。文献[4]在单语、多语和跨语言零样本设置下,对添加了 LSTM 层的 Transformer 的模型进行了评估,结果表明其在英文数据集上可以取得较好的效果。文献[9]在 5 种不同的语言上使用顺序模型来集成包含 LSTM 在内的几个神经网络来学习对于定义的表达。文献[10]采用随机初始化的 mDeBERTa-base 模型对多语言训练数据集进行多任务预训练。经过掩码语言建模和对比学习两个阶段的预训练后,再针对反向字典任务进行微调。

为了提升反向词典模型的性能,有些研究将反向词典任务和其他相关的自然语言处理任务进行联合训练。文献

[11-14]将定义建模任务与反向字典任务共同训练,以期获得更好的定义表示。文献[15]面向基于特征的嵌入,通过组合来自多种语言的训练数据来训练多语言模型;还使用额外的俄语和法语的训练数据,使用无监督嵌入对齐和机器翻译来改进词语解释匹配。

还有一些研究从多个语义层面捕捉定义语义,文献[16]受到文献[17]的启发,多尺度地考虑了定义的不同语义层次的信息,进而匹配合适的目标单词。本研究正是在这项研究的基础上进行的。

2.2 对比学习

反向词典是一种把定义映射到目标词语语义空间的回归问题。传统的回归问题的方法都是只在正样本上进行训练的。如果引入负样本,则可以让定义和目标词语在语义空间中更加靠近,而和其他单词更加远离。这种把对比学习应用到反向词典的方法无疑可以提高反向词典的准确性。

若要利用对比学习进行反向词典构建,则需要解决一个重要的问题,即如何构建负例样本。有监督对比学习的核心就是通过构建负样本实现数据增强,从而改进模型的性能。文献[18]使用一种基于标签信息的负采样策略,将与锚点具有不同标签的样本作为负样本。为了动态地调整负样本的难度,同时提高模型的鲁棒性,文献[19]使用一种基于模型不确定性的负采样策略,根据模型对锚点和候选负样本的预测概率,选择最有可能被误分类的负样本。由此可以看出,针对不同的问题需要构建不同的负样本。

3 研究方法

为了自动生成高质量的提示语句,本文提出了一种基于对比学习的提示语句生成方法。该方法分为两个步骤:首先利用 RDMCL 生成候选词表,然后根据词表生成一个高质量的提示语句。

3.1 RDMCL 基于对比学习的反向词典模型

RDMCL 的目标是从多个语义层面理解定义,从而给出一个小范围的候选词表,其整体架构如图 2 所示。利用 RDMCL 可以给出一个候选单词列表 $Dic = \{word_1, word_2, \dots, word_n\}$,使用 Dic 可以获得高质量的提示语句 P 。

3.1.1 句子级别的语义信息

本文认为句子级别的语义信息应包含句子的整体信息和顺序信息。为了充分学习定义的双向长期依赖关系,选择采用基于 RNN^[20] 的 Bi-LSTM^[21]。Bi-LSTM 由前向 $LSTM_L$ 和后向 $LSTM_R$ 组成。Bi-LSTM 以 emb_s 为输入,按照正向和反向的顺序分为 \vec{emb}_s 和 \overleftarrow{emb}_s 。通过 $LSTM_L$ 将 \vec{emb}_s 转化为 \vec{E} ,通过 $LSTM_R$ 将 \overleftarrow{emb}_s 转化为 \overleftarrow{E} 。将 \vec{E} 和 \overleftarrow{E} 连接成 E_{emb_s} 作为最终结果。其数学描述如下:

$$BiLSTM(e_s) = LSTM_L(\vec{e}_s) + LSTM_R(\overleftarrow{e}_s) \quad (1)$$

$$\vec{E}, \vec{h} = LSTM_L(\vec{emb}_s) \quad (2)$$

$$\overleftarrow{E}, \overleftarrow{h} = LSTM_R(\overleftarrow{emb}_s) \quad (3)$$

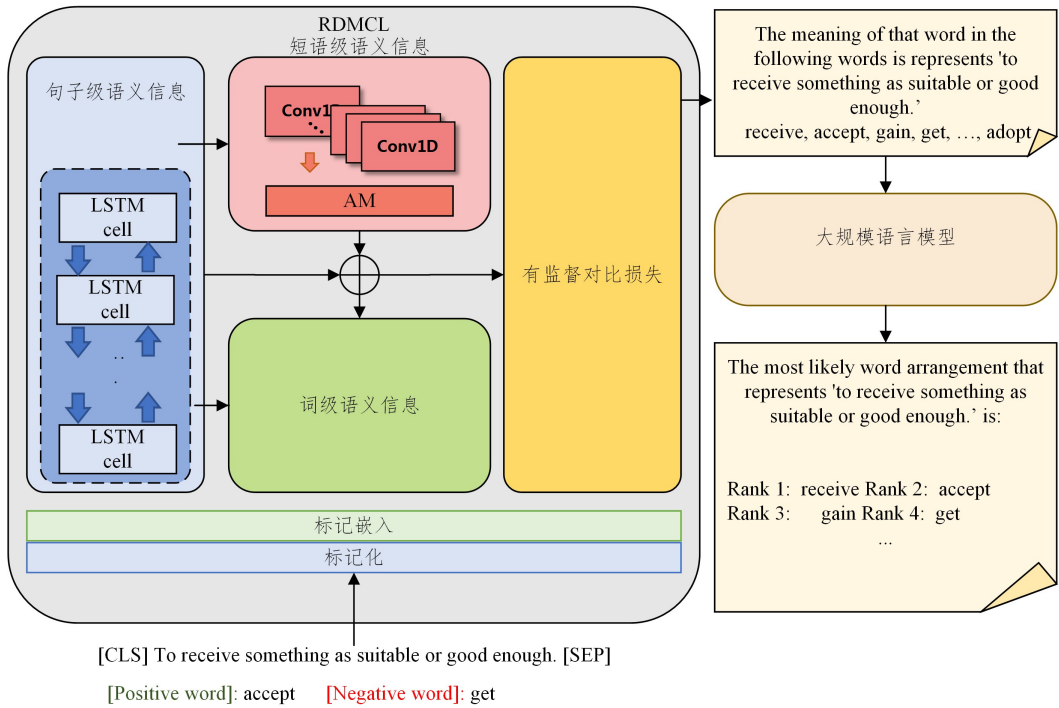


图 2 RDMCL 结构图

Fig. 2 RDMCL structure diagram

$$E_{emb_s} = \vec{E} + \overleftarrow{E} \quad (4)$$

$$V_{sen} = \vec{h} + \overleftarrow{h} \quad (5)$$

其中, \vec{h} 和 \overleftarrow{h} 表示最后一个时间步的隐藏状态, V_{sen} 表示两个方向上隐藏状态的组合结果。

3.1.2 短语级别的语义信息

不仅需要从句子上理解定义的语义信息,还需要从短语级别上进行考虑。短语级别的语义信息通常由几个具有独特语义的词组成,它代表定义的局部特征。卷积神经网络在提取局部特征方面具有良好的表现^[22]。本文通过将卷积

神经网络与注意力机制相结合的方式,进一步增强了模型提取局部特征的能力。使用卷积注意力网络将定义的句子级信息 E_{emb_s} 转换为 V_{phrase} 。其数学描述如下:

$$c_{r,k} = \tanh\left(\sum_{i,j} (F_r * E_{emb_s}^{k:k+l-1})_{i,j}\right) \quad (6)$$

$$c_r = (c_{r,1}, c_{r,2}, \dots, c_{r,t+l+1}) \quad (7)$$

$$C = (c_1, c_2, \dots, c_t) \quad (8)$$

$$i = \text{softmax}(\tanh(W_1 C + b_1)) \quad (9)$$

$$V_{phrase} = (I * E_{emb_s}^T) \quad (10)$$

其中采用的注意力机制类似于自我注意力机制^[23]。 $E_{emb_s}^{k:k+l-1}$ 表示 E_{emb_s} 中第 k 个词向量到第 $(k+l-1)$ 个词向量的矩阵。 F_r 是窗口大小为 l 的卷积核,共有 r 个。式(6)~式(8)是对 E_{emb_s} 的卷积运算, C 是卷积运算后的特征矩阵。特征矩阵 C 包含了句子 E_{emb_s} 的短语级语义信息。 W_1 表示一个可训练的参数矩阵, b_1 表示一个可训练的偏置向量。根据式(9)进行操作后,得到注意力向量 i 。 I 是 i 广播操作后得到的矩阵,它的每个维度的值为该维度所对应词向量的重要性。 V_{phrase} 表示对矩阵行求和生成的特征向量,有效融合 C 中的短语级语义信息。

3.1.3 单词级别的语义信息

除了句子和短语级别的语义信息外,单词级别的语义信息也是理解句子的关键。采用与短语级语义信息获取方法中结构相同但参数不同的注意机制来放大关键词的影响并减小虚拟词的影响。其数学描述如下:

$$i' = \text{softmax}(\tanh(W_2 E_{emb_s}^T + b_2)) \quad (11)$$

$$V_{word} = \text{sum}(I' * E_{emb_s}^T) \quad (12)$$

其中, W_2 表示一个可训练的参数矩阵, b_2 表示一个可训练偏置。特征向量 V_{word} 包含句子 E_{emb_s} 的词级语义信息。

3.1.4 有监督对比学习损失函数

本文将代表定义的不同语义层次的 3 个特征向量 V_{sen} , V_{phrase} , V_{word} 融合为一个特征向量 V_{def} 。使用 V_{def} 和目标词向量 V_{tar} 进行损失函数计算来训练 RDMCL。反向词典模型的效果在一定程度上依赖于目标词向量的质量。 V_{tar} 质量越高,模型的效果越好^[17,24]。为了获得带有语境信息的表示向量,使用冻结参数 Bert-base^[25] 模型将带有目标词的上下文进行嵌入表示。然后,从中选择对应目标单词的语义表示向量作为 V_{tar} 。

同时,为了将负例样本引入模型中,提升模型的泛化能力,本文提出了一种全新的有监督对比学习损失函数 SCL ,该损失函数用于计算损失值,从而训练 RDMCL。其数学描述如下:

$$SCL = \frac{\sum_{i=0}^{n_p-1} \|V_i^{def} - V_i^{Pos_{tar}}\|^2}{n_p} + \frac{n_N T}{\sum_{j=0}^{n_N-1} \|V_j^{def} - V_j^{Neg_{tar}}\|^2} \quad (13)$$

其中, n_p 代表正例样本个数, V^{def} 代表定义的语义表示向量, $V^{Pos_{tar}}$ 代表定义所对应目标单词的语义表示向量, n_N 代表负例样本个数, $V^{Neg_{tar}}$ 代表负例单词的语义表示向量, T 代表温度参数。训练目标是使得 $\|V_i^{def} - V_i^{Pos_{tar}}\|^2$ 尽可能大, $\|V_j^{def} - V_j^{Neg_{tar}}\|^2$ 尽可能小,从而最小化对比损失函数。

3.2 提示语句构造

利用 RDMCL 可以给出 n 个最有可能代表输入的定义

语义的单词。这些单词可以构成一个候选单词列表 $Dic = \{word_1, word_2, \dots, word_n\}$ 。将列表的内容与模板语句 L 进行拼接,获取提示语句 P 。将 P 送入大型语言模型中,期望获得更出色的结果。

4 实验结果与分析

4.1 数据集

Oxford 词典数据集^[7] 中的每条数据包含单词 W 、定义 D 、上下文 C ,每条数据可以由一个三元组 (W, D, C) 表示。然而,该数据集中存在着一些固有问题,例如上下文 C 中不包含单词 W 。本文重新构建了一个新的反向词典数据集。首先将上下文中不包含目标单词的数据条目清洗掉,然后使用清洗后的训练数据集作为正例样本训练集 DP 。以不重复的方式在 DP 中随机选择一个 W^* 替换三元组 (W, D, C) 中的 W ,构成一个新的三元组 (W^*, D, C) ,进而构成一个负例样本训练数据集 DN 。将正例样本和负例样本训练数据集中的数据随机散布在一起构成训练集。

为了更精确地评估模型性能,采用了文献[16]提出的把回归任务转化为分类任务的评价方法。本文从清洗过的 Oxford 数据集的验证集和测试集中随机选择 3000 个数据作为正例样本测试集。利用构建负例样本训练集的方法,从剩余的数据中随机选取 W^* 构造 3000 个负例测试样本,进而构建出整体的验证集和测试集。对于数据集的统计情况如表 1 所列。

表 1 数据集统计

名称	训练集	验证集	测试集
Oxford	97855	12232	12232
Revise Oxford	85455	10784	10666
New Reverse Dictionary	170910	6000	6000

4.2 实验细节和参数设置

本文采用了两种不同的标注和嵌入方法。一种是 GLOVE^[26],它与基准模型使用相同的嵌入方式,以突出 RDMCL 架构的优势。另一种是 BERT-base,它可以进一步提高 RDMCL 的效果。

这些模型使用 AdamW 优化器^[27] 实现了最佳性能。批次大小为 64,初始学习率设置为 0.001,更新 20 个 epoch 的参数,并选择验证损失最低的检查点。

所有的实验都是在 Pytorch 深度学习框架上实现的,并使用 GTX 2080TI GPU 进行训练。

RDMCL 的注意机制网络部分是自适应参数,因此不需要设置超参数。表 2 列出了其他的超参数。

表 2 网络的超参数

名称	单元数
BiLSTM 隐藏层参数	768
卷积和池化层数	2
卷积核数和窗口大小	768,3
池化类型和窗口大小	最大池化,2
温度参数 T	0.0001

4.3 评价指标

本文采用了两种不同类别的评价指标,对模型性能进行了全面准确的评估。一种是分类指标,由文献[16]提出,将回归任务转化为分类任务进行评价。具体地,设定分类阈值 T ,其值为正例样本测试集中计算得到的语义表示与目标词向量余弦相似度的平均值 PA 和负例样本测试集中计算得到的余弦相似度平均值 NA 的均值。然后,根据语义表示与目标词向量余弦相似度是否达到或超过 T 来判定单词是否符合语义。如果余弦相似度不低于 T ,则认为是符合的;否则认为是不符合的。最后利用分类任务中常用的 Precision(P), Recall(R), F1 这 3 个指标对模型进行评价。

表 3 不同方法的表现

Table 3 Performance of different methods

Model	Classification metrics					Regression metrics		
	P \uparrow	R \uparrow	F1 \uparrow	PA \uparrow	NA \downarrow	D \uparrow	Top1 \uparrow	Top5 \uparrow
FCNN ^[17]	65.6	75.4	70.2	24.8	10.5	14.2	10.8	22.2
CNN ^[17]	65.0	62.1	63.5	21.0	10.3	10.4	7.7	16.0
LSTM ^[17]	83.3	28.9	42.9	39.3	16.7	22.7	19.9	36.7
CNN-LSTM ^[17]	74.3	68.3	71.1	24.9	8.9	15.9	16.4	30.6
TransExplain ^[16]	74.3	70.2	72.2	36.8	15.7	21.1	20.3	37.3
Ours								
RDMCL(GLOVE)	80.3	77.8	79.0	41.5	14.9	26.6	25.6	45.4
RDMCL	84.3	80.1	82.1	56.5	37.3	19.2	40.3	64.2

实验 2 为了探究对比学习的有效性,本文开展了相应的消融实验。本文进行了两组消融实验:1) WCL(Without-ContrastiveLearning):不使用对比学习的 RDMCL;2) RD-MCL:使用完整的 RDMCL。实验结果如图 3 所示。

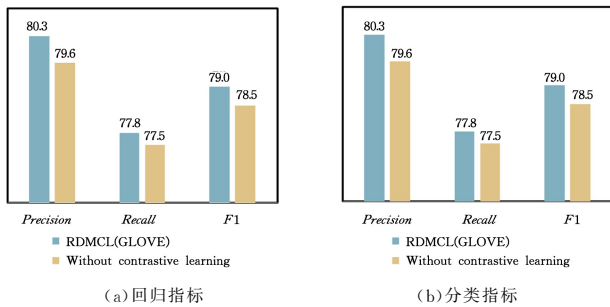


图 3 消融实验结果

Fig. 3 Ablation experimental results

实验 3 本文的目的是利用 RDMCL 生成小规模候选词集,以便创建高质量的提示语句。因此,候选词集是否包含目标单词非常关键。本文研究了候选词集的不同规模对包含目标单词概率的影响。实验结果如图 4 所示。

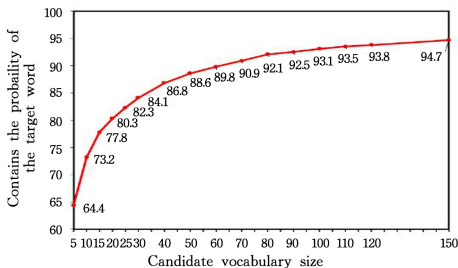


图 4 候选单词集大小和包含目标单词概率的关系图
Fig. 4 Diagram of the relationship between candidate vocabulary size and probability of containing target words

本文还采用了另一种评价方法,即对结果排名进行评估。由于模型输入长度受到限制,无法将整个数据集中的所有单词送入模型进行排序,因此无法使用文献[1]中提出的 rank 值来评估模型。为此,本文提出了类似的评价指标 Top1 和 Top5 值。Top1 和 Top5 值表示模型给出的第一个和前五个最有可能表达定义语义的单词中包含目标单词的概率。

4.4 实验结果

本文在新反向词典数据集上设置了 5 组实验。

实验 1 为了验证所提 RDMCL 方法的有效性和先进性,本文将所提模型与包括 TransExplain 在内的多种可以用于反向词典任务的方法进行了对比,对比实验结果如表 3 所列。

实验 4 本文针对大规模语言模型在使用候选词表时的回答结果进行了优化。具体而言;如果大规模语言模型的回答包含不在候选词表中的单词,则强调候选词表的重要性,并让其重新回答;如果大规模语言模型的回答完全由候选词表中的单词构成,则只需提醒其候选词表的高质量 and 重要性,并让其修改回答。为了探究该方法的有效性,开展了相应的对比实验,结果如图 5 所示。受实验设备的限制,暂时只对比了下列 4 类。1)GPT3;GPT3^[28]是一个由 OpenAI 研发,具有 175B 参数规模的大型语言模型。本课题使用开源的 API 接口测试了 GPT3 模型执行反向词典任务的能力。2)ChatGLM-6B;ChatGLM-6B 是清华大学基于 GLM^[29] 开发的 6B 参数规模的会话模型。本课题测试 ChatGLM-6B 执行反向词典任务的能力。3)GPT3-Prompt:测试采用 RDMCL 构造的高质量提示语句后,GPT3 模型展现出的能力。4)ChatGLM-6B-Prompt:测试采用 RDMCL 构造的高质量提示语句后,ChatGLM-6B 模型展现出的能力。

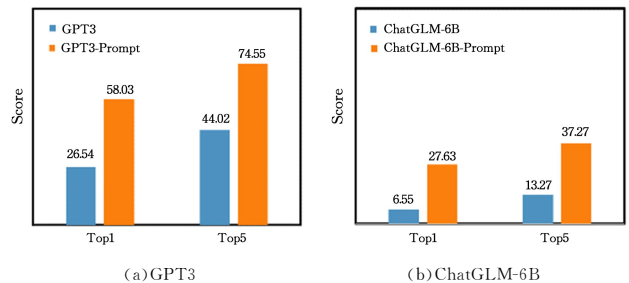


图 5 对比实验结果

Fig. 5 Results of comparative experiments

实验 5 案例学习。为了展示对大规模语言模型给出结果的优化细节,本文在 ChatGLM-6B^[29] 和 GPT3^[28] 模型上分别进行了展示,结果如表 4 所列。

表4 大规模语言模型案例学习

Table 4 Case studies on the performance of large-scale language models

输入定义: (In fantasy literature and games) a member of an imaginary race of human-like creatures, characterized as ugly, warlike, and malevolent.	案例 1(ChatGLM-6B) 问:Please only give me five English words closest to the meaning of '(In fantasy literature and games) a member of an imaginary race of human-like creatures, characterized as ugly, warlike, and malevolent' among the following words {orc, hedgehog, agouti, walrus, peccary, dachshund, ...}	输入定义: A compartment on a ship between decks; often used as a hospital.	案例 2(GPT3) 问:Please only give me five English words closest to the meaning of 'a compartment on a ship between decks; often used as a hospital among' in the following words: {bay, jetty, berth, bay, mouth, run, stateroom, lighter, outdoors, ...} 答:berth stateroom platform sanctum davit
目标单词: orc	答:agoutiorc gobshite dachshundpet 问:The 'gobshite' in the result you provided is not in the candidate word list I provided. Please focus on my candidate word list and provide a new answer. 答:orc dachshund spaniel cowhide savannah	目标单词: bay	问:The ranking probability of this candidate vocabulary represents the degree to which it conforms to the defined semantics, which means that the first word is the most likely to express the defined semantics. After considering this factor, please provide the five most suitable ones 答:berth stateroom bay cabin ward

4.5 实验结果分析

在实验 1 中,FCNN 可以被看作是只考虑单词级别语义信息的网络,CNN 是只使用短语级别语义信息的网络,LSTM 是只使用句子级别语义信息的网络。由表 3 可知,这些网络在性能上均表现不佳,说明只从某一种语义尺度上考虑定义的语义信息并不能取得良好的效果。CNN-LSTM 从短语和句子尺度上考虑定义的语义信息,TransExplain 从单词、短语、句子 3 种不同的层次上考虑定义的语义信息。它们的效果是明显优于从单一尺度上考虑定义语义模型的性能的。但是,它们的性能弱于 RDMCL(GLOVE),这是因为基线模型都是单独考虑某个语义层级上的语义信息的。而 RDMCL 模型在句子层次信息的基础上考虑了词级和短语级语义信息,结果表明这种做法可以在关键指标 Top1 值上提升 5.3%,在 Top5 值上提升 8.1%。

在实验 2 中,本文进行了相关的消融实验以证明对比学习的有效性。由图 3 可知,没有使用对比学习的 RDMCL 在关键指标 Top5 值上下降了 0.7%。这充分表明了对比学习可以增强模型的泛化能力,使得定义的语义表示向量在距离目标单词更近的同时,能够放大它与那些与其语义相悖的单词之间的距离。

在实验 3 中,本文研究了候选词集的大小对包含目标单词概率的影响。从图 4 可以看出,当候选词集的规模为 70 个单词时,包含目标单词的概率就超过了 90%。当规模为 150 个单词时,概率更是达到了 95%。因此,本文选择 150 个单词作为候选词集的合适规模。

在实验 4 中,本文验证了 RDMCL 能够有效地提高提示语句的质量。从图 5 可以看出,高质量的提示语句在两种不同的大模型上都有显著的效果。这充分表明了提示语句的质量对于大模型给出答案的质量有着重要的影响。

实验 5 展示了如何优化大规模语言模型在使用候选词表时的回答结果。根据回答质量的不同,自动采用相应的提示语句。从表 4 的案例可以看出,该方法可以改善大规模语言模型在反向词典任务上的表现。

结束语 本文提出了一种基于对比学习的大模型反向词典任务提示生成方法。该方法通过 RDMCL 从多个语义层面上考虑定义的语义信息,并引入负例样本的对比学习方法增强了模型的泛化能力。RDMCL 可以给出一个小范围的候选

单词列表,进而生成更高质量的提示语句,激发了大模型执行反向词典任务的能力。然而,该方法依然存在一些不足。例如,只能针对英文的反向词典任务,未来可以尝试迁移至多种语言。实验结果表明,改进提示语句的质量对于激发大模型的能力是最简单有效的方法,在未来的研究中,可以尝试其他的改进提示语句质量的方法。此外,RDMCL 的本质是一个语义表示模型,因此可以尝试将其迁移至其他自然语言处理任务上。例如,在情感分析、关系抽取、图文匹配等任务中对文本信息的语义表示就可以延用 RDMCL 的方法。这些都会是未来的工作重点。

参考文献

- [1] MICKUS T, VAN DEEMTER K, CONSTANT M, et al. SemEval-2022 Task 1: CODWOE-Comparing Dictionaries and Word Embeddings[C]// Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022). 2022:1-14.
- [2] QI F, ZHANG L, YANG Y, et al. Wantwords: An open-source online reverse dictionary system[C]// Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: system demonstrations, 2020:175-181.
- [3] ARDOIZ A, ORTEGA-MARTIN M, GARCIA-SIERRA O, et al. MMG at SemEval-2022 Task 1: A Reverse Dictionary approach based on a review of the dataset from a lexicographic perspective[C]// Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022). 2022:68-74.
- [4] TRANT H H, MARTIN M, PURVER M, et al. JSI at SemEval-2022 Task 1: CODWOE-Reverse Dictionary: Monolingual and cross-lingual approaches[C]// Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022). 2022:101-106.
- [5] KHOSLA P, TETERWAK P, WANG C, et al. Supervised contrastive learning [J]. Neural Information Processing Systems, 2020, 33:18661-18673.
- [6] YE H, ZHANG N, DENG S, et al. Contrastive triple extraction with generative transformer[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2021:14257-14265.
- [7] GADETSKY A, YAKUBOVSKIY I, VETROV D. Conditional Generators of Words Definitions[C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics.

- tics (Volume 2; Short Papers). 2018; 266-271.
- [8] CHANG T Y, CHI T C, TSAI S C, et al. xSense: Learning sense-separated sparse representations and textual definitions for explainable word sense networks [J]. arXiv: 1809. 03348, 2018.
- [9] BENDAHDAN N, BRETON J, NICOLAIEFF L, et al. Research at SemEval-2022 Task 1: Deep networks for Reverse Dictionary using embeddings and LSTM autoencoders [C] // Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022). 2022; 94-100.
- [10] LI B, WENG Y, XIA F, et al. LingJing at SemEval-2022 task 1: Multi-task self-supervised pre-training for multilingual reverse dictionary [C] // Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022). 2022; 29-35.
- [11] CHEN P, ZHAO Z. Edinburgh at SemEval-2022 Task 1: Jointly Fishing for Word Embeddings and Definitions [C] // Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022). 2022; 75-81.
- [12] KORENČIĆ D, GRUBISIC I. IRB-NLP at SemEval-2022 Task 1: Exploring the Relationship Between Words and Their Semantic Representations [C] // Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022). 2022; 36-59.
- [13] KONG C, WANG Y, CHONG R, et al. BLCU-ICALL at SemEval-2022 Task 1: Cross-Attention Multitasking Framework for Definition Modeling [C] // Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022). 2022; 23-28.
- [14] SRIVASTAVA A, VEMULAPATI H V. TLDR at SemEval-2022 task 1: Using transformers to learn dictionaries and representations [C] // Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022). 2022; 60-67.
- [15] CERNAVSKI R, STYMNE S. Uppsala University at SemEval-2022 Task 1: Can Foreign Entries Enhance an English Reverse Dictionary? [C] // Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022). 2022; 88-93.
- [16] LI R, LI Z, HUANG S, et al. TransExplain: Using neural networks to find suitable explanations for Chinese phrases [J]. Expert Systems with Applications, 2021, 183: 115440.
- [17] CHANG T Y, CHEN Y N. What does this word mean? explaining contextualized embeddings with natural language definition [C] // Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). 2019; 6064-6070.
- [18] HAFIDI H, GHOGHO M, CIBLAT P, et al. Negative sampling strategies for contrastive self-supervised learning of graph representations [J]. Signal Processing, 2022, 190: 108310.
- [19] JIANG R, NGUYEN T, ISHWAR P, et al. Supervised Contrastive Learning with Hard Negative Samples [J]. arXiv: 2209. 00078, 2022.
- [20] HOPFIELD J J. Neural networks and physical systems with emergent collective computational abilities [J]. Proceedings of the National Academy of Sciences, 1982, 79(8): 2554-2558.
- [21] AUGUSTYNIAK Ł, KAJDANOWICZ T, KAZIENKO P. Aspect detection using word and char embeddings with (Bi) LSTM and CRF [C] // 2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE). IEEE, 2019; 43-50.
- [22] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition [J]. IEEE, 1998, 86: 2278-2324.
- [23] LIN Z, FENG M, DOS SANTOS C, et al. A structured self-attentive sentence embedding [C] // International Conference on Learning Representations. International Conference on Learning Representations. ICLR, 2017.
- [24] NORASET T, LIANG C, BIRNBAUM L, et al. Definition modeling: Learning to define word embeddings in natural language [C] // Proceedings of the AAAI Conference on Artificial Intelligence. 2017.
- [25] KENTON J D M-W C, TOUTANOVA L K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding [C] // Proceedings of NAACL-HLT. 2019; 4171-4186.
- [26] PENNINGTON J, SOCHER R, MANNING C D. Glove: Global vectors for word representation [C] // Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014; 1532-1543.
- [27] LOSHCHELOV I, HUTTER F. Decoupled Weight Decay Regularization [C] // International Conference on Learning Representations. 2017
- [28] BROWN T, MANN B, RYDER N, et al. Language models are few-shot learners [J]. Neural Information Processing Systems, 2020, 33: 1877-1901.
- [29] ZENG A, LIU X, DU Z, et al. Glm-130b: An open bilingual pre-trained model [J]. arXiv: 2210. 02414, 2022.



TIAN Sicheng, born in 1997, Ph.D. His main research interests include natural language processing and smart health-care.



HUANG Shaobin, born in 1965, Ph.D., professor, Ph.D supervisor. His main research interests include machine learning and natural language processing.