

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2024 年 2 月 15 日 (15.02.2024)



(10) 国际公布号
WO 2024/031891 A1

- (51) 国际专利分类号:
G06F 40/30 (2020.01) **G06F 16/33** (2019.01)
- (21) 国际申请号: PCT/CN2022/137938
- (22) 国际申请日: 2022 年 12 月 9 日 (09.12.2022)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202210955108.0 2022年8月10日 (10.08.2022) CN
- (71) 申请人: 浙江大学 (ZHEJIANG UNIVERSITY) [CN/CN]; 中国浙江省杭州市西湖区余杭塘路 866 号, Zhejiang 310058 (CN)。
- (72) 发明人: 张宁豫 (ZHANG, Ningyu); 中国浙江省杭州市西湖区浙大路 38 号, Zhejiang 310027 (CN)。李磊 (LI, Lei); 中国浙江省宁波市鄞

州区江南路 1689 号, Zhejiang 315100 (CN)。陈想 (CHEN, Xiang); 中国浙江省杭州市西湖区浙大路 38 号, Zhejiang 310027 (CN)。陈华钧 (CHEN, Huajun); 中国浙江省杭州市西湖区浙大路 38 号, Zhejiang 310027 (CN)。

- (74) 代理人: 杭州天勤知识产权代理有限公司 (HANGZHOU TIANQIN INTELLECTUAL PROPERTY AGENCY CO., LTD.); 中国浙江省杭州市西湖区竞舟路 1 号筑品金座 501 室, Zhejiang 310013 (CN)。
- (81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ,

(54) **Title:** FINE TUNING METHOD AND APPARATUS FOR KNOWLEDGE REPRESENTATION-DISENTANGLED CLASSIFICATION MODEL, AND APPLICATION

(54) 发明名称: 知识表征解耦的分类模型的微调方法、装置和应用

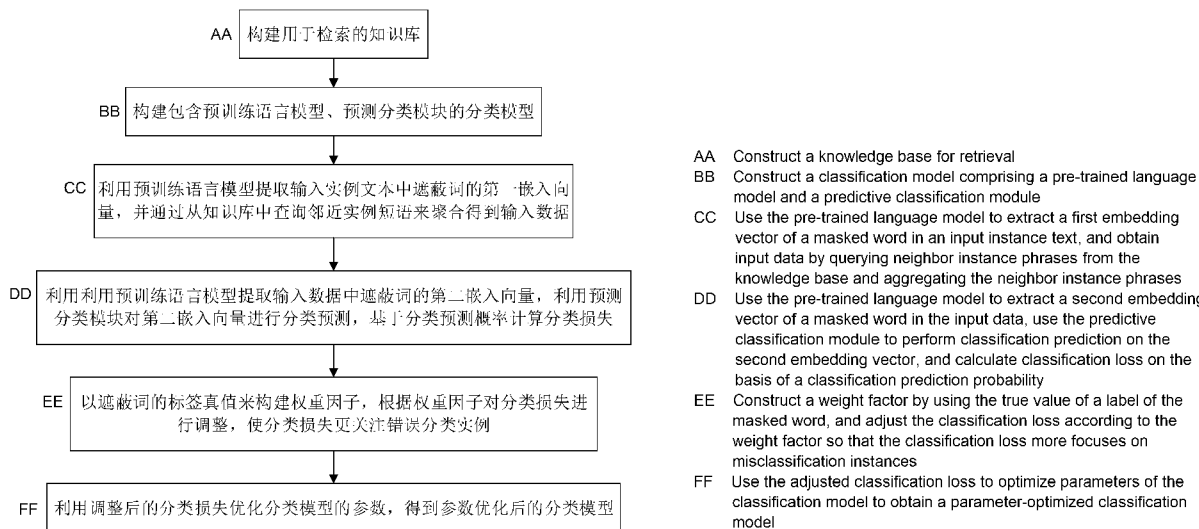


图 1

(57) **Abstract:** Disclosed are a fine tuning method and apparatus for a knowledge representation-disentangled classification model, and an application. A knowledge representation and a classification model are disentangled and stored in a knowledge base, and are matched and aggregated according to retrieval during application, so that rote memorization of a learning model is restricted and the generalization ability of the model is improved. Moreover, neighbor instance phrases are retrieved from the knowledge base by means of KNN as consecutive neural examples, and the neural examples are utilized to guide classification model training and correct classification model prediction, so that the ability of the classification model in few-sample and zero-sample scenarios is improved. When the data volume is large enough, the knowledge base has better and richer information accordingly, and the classification model also has prominent performance in a fully supervised scenario.



LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:
— 包括国际检索报告(条约第21条(3))。

(57) 摘要: 本发明公开了一种知识表征解耦的分类模型的微调方法、装置和应用, 将知识表征与分类模型解耦, 存储于知识库中, 应用的时候根据检索进行匹配聚合, 这样限制了学习模型的死记硬背, 提高了模型的泛化能力, 同时利用KNN从知识库中检索得到邻近实例短语作为连续的神经示例, 利用神经示例指导分类模型训练和纠正分类模型预测, 提高了分类模型在少样本和零样本场景下的能力, 当数据量足够多时, 知识库相应也拥有更佳更丰富的信息, 分类模型在全监督场景下表现也十分突出。

知识表征解耦的分类模型的微调方法、装置和应用

技术领域

本发明属于自然语言处理技术领域，具体涉及一种知识表征解耦的分类模型的微调方法、装置和应用。

背景技术

预训练分类模型通过从海量数据中深度学习知识，在自然语言处理领域取得了激动人心的显著成果。预训练分类模型通过设计通用的预训练任务，如遮蔽掩码建模（MLM）、下句预测（NSP）等，从大规模的语料中进行训练，在应用到下游关系分类、情感分类等分类任务时，只需使用少量数据微调预训练分类模型，便能取得良好性能。

提示学习的出现，减少了预训练分类模型在微调阶段与预训练阶段的差异性，使得预训练分类模型进一步具备了少样本和零样本学习的能力。提示学习可分为离散提示和连续提示，离散提示通过人工构建离散的提示模板来转换输入形式，连续提示在输入序列中添加一系列可学习的连续嵌入向量，减少了提示工程。

然而，最近的研究表明当数据量及其匮乏时，预训练分类模型的泛化能力不尽人意。一个潜在的原因在于，参数化模型通过记忆的方式很难掌握稀疏和困难样本，导致不充分的泛化能力。当数据呈现长尾分布并且具有小的非典型实例集群，预训练分类模型倾向于通过死记硬背这些非典型实例而不是通过学习更通用的模式知识来进行预测，这会导致预训练分类模型学习的知识表示在下游分类任务中表现差，分类结果准确率不高。

专利文献 CN101127042A 公开了一种基于分类模型的情感分类方法，

专利文献 CN108363753A 公开了一种评论文本情感分类模型训练与情感分类方法、装置及设备，这两篇专利申请均是通过提取文本的嵌入向量后，基于嵌入向量来构建进行情感分类。这两种方式当样本数据匮乏时，由于提取的嵌入向量不佳，就难实现情感分类的准确性。

发明内容

针对现有技术所存在的上述技术问题，本发明的目的是提供一种知识表征解耦的分类模型的微调方法、装置和应用，通过将分类模型得到的知识表征解耦成知识库，该知识库作为相似度引导来优化分类模型，以提高分类模型知识表示的能力和准确性，进而提高下游分类任务的分类准确性。

为实现上述发明目的，实施例提供的一种知识表征解耦的分类模型的微调方法，包括以下步骤：

步骤 1，构建用于检索的知识库，知识库中存有多个实例短语，每个实例短语以键值对的形式存储，其中键存储实例词语的嵌入向量，值存储实例短语的标签真值；

步骤 2，构建包含预训练语言模型、预测分类模块的分类模型；

步骤 3，利用预训练语言模型提取输入实例文本中遮蔽词的第一嵌入向量，并以该第一嵌入向量作为第一查询向量，针对每个标签类别从知识库中查询与第一查询向量最邻近的多个实例短语作为第一邻近实例短语，将所有第一邻近实例短语与第一查询向量聚合得到的聚合结果作为预训练语言模型的输入数据；

步骤 4，利用预训练语言模型提取输入数据中遮蔽词的第二嵌入向量，利用预测分类模块对第二嵌入向量进行分类预测，以得到分类预测概率，基于该分类预测概率和遮蔽词的标签真值计算分类损失；

步骤 5，以遮蔽词的标签真值来构建权重因子，根据权重因子对分类损失进行调整，使分类损失更关注错误分类实例；

步骤 6，利用调整后的分类损失优化分类模型的参数，得到参数优化后的分类模型。

为实现上述发明目的，实施例提供的一种知识表征解耦的分类模型的微调装置，包括：

知识库构建和更新单元，用于构建用于检索的知识库，知识库中存多个实例短语，每个实例短语以键值对的形式存储，其中键存储实例词语的嵌入向量，值存储实例短语的标签真值；

分类模型构建单元，用于构建包含预训练语言模型、预测分类模块的分类模型；

查询及聚合单元，用于利用预训练语言模型提取输入实例文本中遮蔽词的第一嵌入向量，并以该第一嵌入向量作为第一查询向量，针对每个标签类别从知识库中查询与第一查询向量最邻近的多个实例短语作为第一邻近实例短语，将所有第一邻近实例短语与第一查询向量聚合得到的聚合结果作为预训练语言模型的输入数据；

损失计算单元，用于利用预训练语言模型提取输入数据中遮蔽词的第二嵌入向量，利用预测分类模块对第二嵌入向量进行分类预测，以得到分类预测概率，基于该分类预测概率和遮蔽词的标签真值计算分类损失；

损失调整单元，用于以遮蔽词的标签真值来构建权重因子，根据权重因子对分类损失进行调整，使分类损失更关注错误分类实例；

参数优化单元，用于利用调整后的分类损失优化分类模型的参数，得到参数优化后的分类模型。

为实现上述发明目的，实施例还提供了一种利用知识表征解耦的分类

模型的任务分类方法，所述任务分类方法应用上述微调方法构建的知识库和参数优化后的分类模型，包括以下步骤：

步骤 1，利用参数优化后的预训练语言模型提取输入实例文本中遮蔽词的第三嵌入向量，并以该第三嵌入向量作为第三查询向量，针对每个标签类别从知识库中查询与第三查询向量最邻近的多个实例短语作为第三邻近实例短语，将所有第三邻近实例短语与第三查询向量聚合得到的聚合结果作为预训练语言模型的输入数据；

步骤 2，利用参数优化后的预训练语言模型提取输入数据中遮蔽词的第四嵌入向量，针对每类从知识库中查询与第四查询向量最邻近的多个实例文本作为第四邻近实例文本，依据第四查询向量与第四邻近实例文本之间的相似度来计算类别相关概率；

步骤 3，利用参数优化后的预测分类模块对第四嵌入向量进行分类预测，以得到分类预测概率；

步骤 4，以每个类别相关概率和分类预测概率的加权结果作为总分类预测结果。

与现有技术相比，本发明具有的有益效果至少包括：

将知识表征与分类模型解耦，存储于知识库中，应用的时候根据检索进行匹配聚合，这样限制了学习模型的死记硬背，提高了模型的泛化能力，同时利用 KNN 从知识库中检索得到邻近实例短语作为连续的神经示例，利用神经示例指导分类模型训练和纠正分类模型预测，提高了分类模型在少样本和零样本场景下的能力，当数据量足够多时，知识库相应也拥有更佳更丰富的信息，分类模型在全监督场景下表现也十分突出。

附图说明

为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图做简单地介绍，显而易见地，下面描述中的附图仅仅是本发明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动前提下，还可以根据这些附图获得其他附图。

图 1 是实施例提供的知识表征解耦的分类模型的微调方法的流程图；

图 2 是实施例提供的分类模型的结构及训练示意图和知识库更新示意图以及分类预测示意图；

图 3 是实施例提供的利用知识表征解耦的分类模型的任务分类方法的流程图。

具体实施方式

为使本发明的目的、技术方案及优点更加清楚明白，以下结合附图及实施例对本发明进行进一步的详细说明。应当理解，此处所描述的具体实施方式仅仅用以解释本发明，并限定本发明的保护范围。

针对传统的提示学习方法和微调方法不能很好的处理非典型样本，导致分类模型的代表能力不强，进而影响分类任务的预测准确性。现有技术通过死记硬背这些非典型实例而不是通过学习更通用的模式知识来进行预测，会导致模型表征能力差，这与人类通过类比来学习知识相反，人类可以通过联想学习来回忆深层记忆中的相关技能，从而相互加强，从而拥有解决小样本和零样本任务的非凡能力。受此启发，实施例提供了一种知识表征解耦的分类模型的微调方法和装置，以及微调后的分类模型的应用，通过从训练实例文本中构建知识库，将记忆从预训练语言模型中解耦，为模型的训练和预测提供参考知识，提高模型的泛化能力。

图 1 是实施例提供的知识表征解耦的分类模型的微调方法的流程图。

如图 1 所示，实施例提供的知识表征解耦的分类模型的微调方法，包括以下步骤：

步骤 1，构建用于检索的知识库。

实施例中，知识库作为一种额外的参考信息将知识表征从分类模型的部分记忆中解耦出来，主要用于存储从分类模型中结构得到的知识表征，该知识表征以市里短语的形式存在，具体地，每个实例短语以键值对的形式存储，其中，键存储实例词语的嵌入向量，值存储实例短语的标签真值。实例短语的嵌入向量是基于提示模板的实例文本经过预训练语言模型学习得到，具体为实例文本中的掩码位置在预训练语言模型中最后一层输出的隐藏向量。

需要说明的是，知识库可以被自由的被添加、编辑和删除，如图 2 所示，每轮训练时，输入实例文本中遮蔽词的第一嵌入向量及其对应的标签真值组成新实例短语，被异步更新到知识库中。

步骤 2，构建包含预训练语言模型、预测分类模块的分类模型。

如图 2 所示，实施例构建的分类模型包含预训练语言模型，该预训练语言模型用于对输入实例文本进行知识表示，以提取掩码位置的嵌入向量，具体地，输入的实例文本需要经过提示模板序列化转化，提示模板的形式为：[CLS]实例文本[MASK] [SEP]，举例说明为：[CLS]这部电影没有任何意义[MASK] [SEP]，同时将标签真值通过映射函数映射到预训练语言模型的词表空间，得到标签向量。预测分类模块用于对输入的嵌入向量进行分类预测以输出分类预测概率。

步骤 3，利用预训练语言模型提取输入实例文本中遮蔽词的第一嵌入向量，并通过从知识库中查询邻近实例短语来聚合得到输入数据。

实施例中，利用预训练语言模型提取输入实例文本中遮蔽词的第一嵌

入向量，并以该第一嵌入向量作为第一查询向量，针对每个标签类别，采用 KNN（最邻近节点算法）搜索从知识库中查询与第一查询向量最邻近的 m 个实例短语作为第一邻近实例短语，该些第一邻近实例短语作为额外的示例输入，与第一查询向量聚合得到的聚合结果作为预训练语言模型的输入数据，其中聚合公式为：

$$I = e(\hat{x}) \oplus \left[\sum_{i \in [1;m]} (\alpha_i^l h_{ei}^l), e(v^l) \right] \oplus \dots \oplus \left[\sum_{i \in [1;m]} (\alpha_i^L h_{ei}^L), e(v^L) \right]$$

$$\alpha_i^l = \frac{e^{h_q \cdot h_{ei}^l}}{\sum_{i \in [1;m]} e^{h_q \cdot h_{ei}^l}}$$

其中， $e(\hat{x})$ 表示经过提示模板序列化处理的输入实例文本的初始向量， h_q 表示输入实例文本中遮蔽词的第一查询向量， h_{ei}^l 表示第 l 类标签中第 i 个第一邻近实例短语的嵌入向量， m 为第一邻近实例短语总量， α_i^l 表示 h_{ei}^l 的 softmax 值，表示与第一查询向量之间的相关性， $e(v^l)$ 表示第一邻近实例短语的标签真值， L 表示标签总量， I 表示聚合得到的聚合结果，该聚合结果作为输入数据结合了来自于知识库的实例短语，作为上下文增强信息，用于指导分类模型训练和纠正分类模型预测，提高了分类模型在少样本和零样本场景下的能力。

步骤 4，利用利用预训练语言模型提取输入数据中遮蔽词的第二嵌入向量，利用预测分类模块对第二嵌入向量进行分类预测，基于分类预测概率计算分类损失。

实施例中，在构建计算分类损失时，以输入数据对应的分类预测概率和遮蔽词的标签真值的交叉熵作为分类损失 L_{CE} 。

步骤 5，以遮蔽词的标签真值来构建权重因子，根据权重因子对分类

损失进行调整，使分类损失更关注错误分类实例。

实施例中，通过遮蔽词的标签真值来调整分类损失中正确分类和错误分类的权重来使分类模型更佳关注错误分类样本，具体公式如下：

$$L = (1 + \beta F(p_{knn}))L_{CE}$$

其中， L_{CE} 表示分类损失， β 表示调节参数， $F(p_{knn})$ 表示权重因子，表示为 $F(p_{knn}) = -\log(p_{knn})$ ， p_{knn} 表示遮蔽词的标签真值。

步骤 6，利用调整后的分类损失优化分类模型的参数，得到参数优化后的分类模型。

实施例中，利用构建的分类损失来优化分类模型的参数，并在每轮训练时，将输入实例文本的第一嵌入向量来构建实例短语，更新到知识库中。

上述知识表征解耦的分类模型的微调方法微调后的分类模型提升了在少样本和零样本场景下的能力。当数据量足够多时，知识库相应也拥有更佳更丰富的信息，分类模型在全监督场景下表现也十分突出。

基于同样的发明构思，实施例还提供了一种知识表征解耦的分类模型的微调装置，包括：

知识库构建和更新单元，用于构建用于检索的知识库，知识库中存有多实例短语，每个实例短语以键值对的形式存储，其中键存储实例词语的嵌入向量，值存储实例短语的标签真值；

分类模型构建单元，用于构建包含预训练语言模型、预测分类模块的分类模型；

查询及聚合单元，用于利用预训练语言模型提取输入实例文本中遮蔽词的第一嵌入向量，并以该第一嵌入向量作为第一查询向量，针对每个标签类别从知识库中查询与第一查询向量最邻近的多个实例短语作为第一邻近实例短语，将所有第一邻近实例短语与第一查询向量聚合得到的聚合

结果作为预训练语言模型的输入数据；

损失计算单元，用于利用预训练语言模型提取输入数据中遮蔽词的第二嵌入向量，利用预测分类模块对第二嵌入向量进行分类预测，以得到分类预测概率，基于该分类预测概率和遮蔽词的标签真值计算分类损失；

损失调整单元，用于以遮蔽词的标签真值来构建权重因子，根据权重因子对分类损失进行调整，使分类损失更关注错误分类实例；

参数优化单元，用于利用调整后的分类损失优化分类模型的参数，得到参数优化后的分类模型。

需要说明的是，上述实施例提供的知识表征解耦的分类模型的微调装置在进行微调分类模型时，应以上述各功能单元的划分进行举例说明，可以根据需要将上述功能分配由不同的功能单元完成，即在终端或服务器的内部结构划分成不同的功能单元，以完成以上描述的全部或者部分功能。另外，上述实施例提供的知识表征解耦的分类模型的微调装置与知识表征解耦的分类模型的微调方法实施例属于同一构思，其具体实现过程详见知识表征解耦的分类模型的微调方法实施例，这里不再赘述。

基于同样的发明构思，实施例还提供了一种利用知识表征解耦的分类模型的任务分类方法，该任务分类方法应用上述微调方法构建的知识库和参数优化后的分类模型，如图 3 所示，包括以下步骤：

步骤 1，利用参数优化后的预训练语言模型提取输入实例文本中遮蔽词的第三嵌入向量，并通过从知识库中查询邻近实例短语来聚合得到输入数据。

实施例中，利用参数优化后的预训练语言模型提取输入实例文本中遮蔽词的第三嵌入向量，并以该第三嵌入向量作为第三查询向量，针对每个标签类别从知识库中查询与第三查询向量最邻近的多个实例短语作为第

三邻近实例短语，将所有第三邻近实例短语与第三查询向量聚合得到的聚合结果作为预训练语言模型的输入数据。

使用非参数化方法 KNN 从知识库中检索与输入实例文本邻近的实例短语，将 KNN 检索的结果视为容易与困难实例的指示信息，让分类模型在训练时更关注困难样本。

步骤 2，利用参数优化后的预训练语言模型提取输入数据中遮蔽词的第四嵌入向量，并通过从知识库中查询邻近实例短语来计算类别相关概率。

实施例中，利用参数优化后的预训练语言模型提取输入数据中遮蔽词的第四嵌入向量，针对每类，采用 KNN 搜索从知识库中查询与第四查询向量最邻近的多个实例文本作为第四邻近实例文本，依据第四查询向量与第四邻近实例文本之间的相似度来计算类别相关概率，具体地，采用以下公式依据第四查询向量与第四邻近实例文本之间的相似度来计算类别相关概率：

$$P_{KNN}(y_i|q_t) = \sum_{(c_i, y_i) \in N} \exp(d(h_{q_t}, h_{c_i}))$$

其中， $P_{KNN}(y_i|q_t)$ 表示输入实例文本 q_t 的第 i 分类类别的类别相关概率， $d(h_{q_t}, h_{c_i})$ 表示输入实例文本 q_t 的第四查询向量 h_{q_t} 与属于第 i 分类类别 y_i 的实例短语 c_i 的嵌入向量 h_{c_i} 之间的内积距离，作为内积相似度， N 表示知识库。

KNN 是一种非参数化方法，可以非常容易的对输入实例文本做出预测，不需要任何的分类层，因此可以直观的将 KNN 的分类结果（类别相关概率）作为一种先验知识来指导预训练分类模型，使其更加关注难样本（或非典型样本）。

步骤 3，利用参数优化后的预测分类模块对第四嵌入向量进行分类预

测，以得到分类预测概率。

步骤 4，以每个类别相关概率和分类预测概率的加权结果作为总分类预测结果。

传统的预训练语言模型在预测时仅依赖模型参数化记忆能力，引入非参数化方法 KNN 后，可以使模型在预测时通过检索最近邻样本来做决策，类似于“开卷考试”。通过 KNN 检索得到类别相关概率 $P_{KNN}(y_i|q_t)$ ，分类模型输出的分类预测概率 $P(y_i|q_t)$ ，将两种概率分布加权求和得到总分类预测结果，表示为：

$$P = \gamma P_{KNN}(y_i|q_t) + (1 - \gamma)P(y_i|q_t)$$

其中， γ 表示权重参数。

通过 KNN 检索得到类别相关概率 $P_{KNN}(y_i|q_t)$ 可以进一步用于分类模型的推理过程，来纠正分类模型在推理时产生的错误。

实施例提供的利用知识表征解耦的分类模型的任务分类方法，可用于关系分类任务。当用于关系分类任务时，知识库中存储的实例短语的标签真值为关系类型，包括朋友关系、亲属关系、同事关系、同学关系，在进行关系分类时，根据输入实例文本经过步骤 1 和 2 计算得到每个关系类型的类别相关概率，根据步骤 3 计算分类预测概率，根据步骤 4 计算每个关系类型对应的总分类预测结果，通过筛选得到最大的总分类预测结果作为输入实例文本对应的最终关系分类结果。

实施例提供的利用知识表征解耦的分类模型的任务分类方法，可用于情感分类任务。当用于情感分类任务时，知识库中存储的实例短语的标签真值为情感类型，包括积极情感、消极情感，在进行情感分类时，根据输入实例文本经过步骤 1 和 2 计算得到每个情感类型的类别相关概率，根据步骤 3 计算分类预测概率，根据步骤 4 计算情感类型对应的总分类预测结

果，通过筛选得到最大的总分类预测结果作为输入实例文本对应的最终情感分类结果。

在情感分类任务中，以 Roberta-large 作为预训练语言模型，为了提高检索速度，使用开源库 FAISS 进行 KNN 检索。输入实例文本为“这部电影没有任何意义!”时，进行情感分类的过程为：

(1) 构建提示模板对输入实例文本进行转换，经过提示模板转换后输入变为 “[CLS]这部电影没有任何意义! [MASK] [SEP]”。

(2) 利用预训练语言模型获得输入实例文本[MASK]位置在嵌入向量，从知识库中检索神经示例，与输入实例文本在[MASK]位置在嵌入向量进行拼接聚合后再输入到预训练语言模型中。

(3) 将输入实例文本[MASK]位置在语言模型最后一层的隐藏状态作为查询向量从知识库中检索最近邻实例短语，基于实例短语计算类别相关概率 $P_{KNN}(y_i|q_t)$ ，其中标签为“差评”概率为 0.8，“好评”的概率为 0.2；

(4) 利用预测分类模块得到查询向量的分类预测概率 $P(y_i|q_t)$ ，其中，标签为“差评”的概率是 0.4，“好评”的概率为 0.6；

(5) 将两种概率 $P_{KNN}(y_i|q_t)$ 和 $P(y_i|q_t)$ 加权求和得到总分类预测结果，权重参数 γ 选择 0.5，这样标签为“差评”的总分类预测概率为 0.6，“好评”的总分类预测概率为 0.4。

以上所述的具体实施方式对本发明的技术方案和有益效果进行了详细说明，应理解的是以上所述仅为本发明的最优选实施例，并不用于限制本发明，凡在本发明的原则范围内所做的任何修改、补充和等同替换等，均应包含在本发明的保护范围之内。

权 利 要 求 书

1、一种知识表征解耦的分类模型的微调方法，其特征在于，包括以下步骤：

步骤 1，构建用于检索的知识库，知识库中存有多实例短语，每个实例短语以键值对的形式存储，其中键存储实例词语的嵌入向量，值存储实例短语的标签真值；

步骤 2，构建包含预训练语言模型、预测分类模块的分类模型；

步骤 3，利用预训练语言模型提取输入实例文本中遮蔽词的第一嵌入向量，并以该第一嵌入向量作为第一查询向量，针对每个标签类别从知识库中查询与第一查询向量最邻近的多个实例短语作为第一邻近实例短语，将所有第一邻近实例短语与第一查询向量聚合得到的聚合结果作为预训练语言模型的输入数据；

步骤 4，利用预训练语言模型提取输入数据中遮蔽词的第二嵌入向量，利用预测分类模块对第二嵌入向量进行分类预测，以得到分类预测概率，基于该分类预测概率和遮蔽词的标签真值计算分类损失；

步骤 5，以遮蔽词的标签真值来构建权重因子，根据权重因子对分类损失进行调整，使分类损失更关注错误分类实例；

步骤 6，利用调整后的分类损失优化分类模型的参数，得到参数优化后的分类模型。

2、根据权利要求 1 所述的知识表征解耦的分类模型的微调方法，其特征在于，步骤 2 中，采用 KNN 检索从知识库中查询与第一查询向量最邻近的多个实例短语作为第一邻近实例短语，并通过以下聚合方式聚合所有第一邻近实例短语与第一查询向量：

$$I = e(\hat{x}) \oplus \left[\sum_{i \in [1;m]} (\alpha_i^l h_{ei}^l), e(v^l) \right] \oplus \cdots \oplus \left[\sum_{i \in [1;m]} (\alpha_i^L h_{ei}^L), e(v^L) \right]$$

$$\alpha_i^l = \frac{e^{h_q \cdot h_{ei}^l}}{\sum_{i \in [1;m]} e^{h_q \cdot h_{ei}^l}}$$

其中, I 表示聚合得到的聚合结果, $e(\hat{x})$ 表示经过提示模板序列化处理的输入实例文本的初始向量, h_q 表示输入实例文本中遮蔽词的第一查询向量, h_{ei}^l 表示第 l 类标签中第 i 个第一邻近实例短语的嵌入向量, m 为第一邻近实例短语总量, α_i^l 表示 h_{ei}^l 的 softmax 值, 表示与第一查询向量之间的相关性, $e(v^l)$ 表示第一邻近实例短语的标签真值, L 表示标签总量。

3、根据权利要求 1 所述的知识表征解耦的分类模型的微调方法, 其特征在于, 步骤 5 中, 调整后的分类损失 L 表示为:

$$L = (1 + \beta F(p_{knn})) L_{CE}$$

其中, L_{CE} 表示分类损失, β 表示调节参数, $F(p_{knn})$ 表示权重因子, 表示为 $F(p_{knn}) = -\log(p_{knn})$, p_{knn} 表示遮蔽词的标签真值。

4、根据权利要求 1 所述的知识表征解耦的分类模型的微调方法, 其特征在于, 包括: 以分类预测概率和遮蔽词的标签真值的交叉熵来计算分类损失。

5、根据权利要求 1-4 任一项所述的知识表征解耦的分类模型的微调方法, 其特征在于, 还包括: 利用预训练语言模型提取的第一嵌入向量及其对应的标签真值形成新实例短语, 更新到知识库中。

6、根据权利要求 1 所述的知识表征解耦的分类模型的微调装置, 其特征在于, 包括:

知识库构建和更新单元, 用于构建用于检索的知识库, 知识库中存有

多个实例短语，每个实例短语以键值对的形式存储，其中键存储实例词语的嵌入向量，值存储实例短语的标签真值；

分类模型构建单元，用于构建包含预训练语言模型、预测分类模块的分类模型；

查询及聚合单元，用于利用预训练语言模型提取输入实例文本中遮蔽词的第一嵌入向量，并以该第一嵌入向量作为第一查询向量，针对每个标签类别从知识库中查询与第一查询向量最邻近的多个实例短语作为第一邻近实例短语，将所有第一邻近实例短语与第一查询向量聚合得到的聚合结果作为预训练语言模型的输入数据；

损失计算单元，用于利用预训练语言模型提取输入数据中遮蔽词的第二嵌入向量，利用预测分类模块对第二嵌入向量进行分类预测，以得到分类预测概率，基于该分类预测概率和遮蔽词的标签真值计算分类损失；

损失调整单元，用于以遮蔽词的标签真值来构建权重因子，根据权重因子对分类损失进行调整，使分类损失更关注错误分类实例；

参数优化单元，用于利用调整后的分类损失优化分类模型的参数，得到参数优化后的分类模型。

7、一种利用知识表征解耦的分类模型的任务分类方法，其特征在于，所述任务分类方法应用权利要求 1-5 任一项所述的微调方法构建的知识库和参数优化后的分类模型，包括以下步骤：

步骤 1，利用参数优化后的预训练语言模型提取输入实例文本中遮蔽词的第三嵌入向量，并以该第三嵌入向量作为第三查询向量，针对每个标签类别从知识库中查询与第三查询向量最邻近的多个实例短语作为第三邻近实例短语，将所有第三邻近实例短语与第三查询向量聚合得到的聚合结果作为预训练语言模型的输入数据；

步骤 2，利用参数优化后的预训练语言模型提取输入数据中遮蔽词的第四嵌入向量，针对每类从知识库中查询与第四查询向量最邻近的多个实例文本作为第四邻近实例文本，依据第四查询向量与第四邻近实例文本之间的相似度来计算类别相关概率；

步骤 3，利用参数优化后的预测分类模块对第四嵌入向量进行分类预测，以得到分类预测概率；

步骤 4，以每个类别相关概率和分类预测概率的加权结果作为总分类预测结果。

8、根据权利要求 7 所述的利用知识表征解耦的分类模型的任务分类方法，其特征在于，采用以下公式依据第四查询向量与第四邻近实例文本之间的相似度来计算类别相关概率：

$$P_{KNN}(y_i|q_t) = \sum_{(c_i, y_i) \in N} \exp(d(h_{q_t}, h_{c_i}))$$

其中， $P_{KNN}(y_i|q_t)$ 表示输入实例文本 q_t 的第 i 分类类别的类别相关概率， $d(h_{q_t}, h_{c_i})$ 表示输入实例文本 q_t 的第四查询向量 h_{q_t} 与属于第 i 分类类别 y_i 的实例短语 c_i 的嵌入向量 h_{c_i} 之间的内积距离，作为内积相似度， N 表示知识库。

9、根据权利要求 7 所述的利用知识表征解耦的分类模型的任务分类方法，其特征在于，当用于关系分类任务时，知识库中存储的实例短语的标签真值为关系类型，包括朋友关系、亲属关系、同事关系、同学关系，在进行关系分类时，根据输入实例文本经过步骤 1 和 2 计算得到每个关系类型的类别相关概率，根据步骤 3 计算分类预测概率，根据步骤 4 计算每个关系类型对应的总分类预测结果，通过筛选得到最大的总分类预测结果作为输入实例文本对应的最终关系分类结果。

10、根据权利要求 7 所述的利用知识表征解耦的分类模型的任务分类方法，其特征在于，当用于情感分类任务时，知识库中存储的实例短语的标签真值为情感类型，包括积极情感、消极情感，在进行情感分类时，根据输入实例文本经过步骤 1 和 2 计算得到每个情感类型的类别相关概率，根据步骤 3 计算分类预测概率，根据步骤 4 计算情感类型对应的总分类预测结果，通过筛选得到最大的总分类预测结果作为输入实例文本对应的最终情感分类结果。

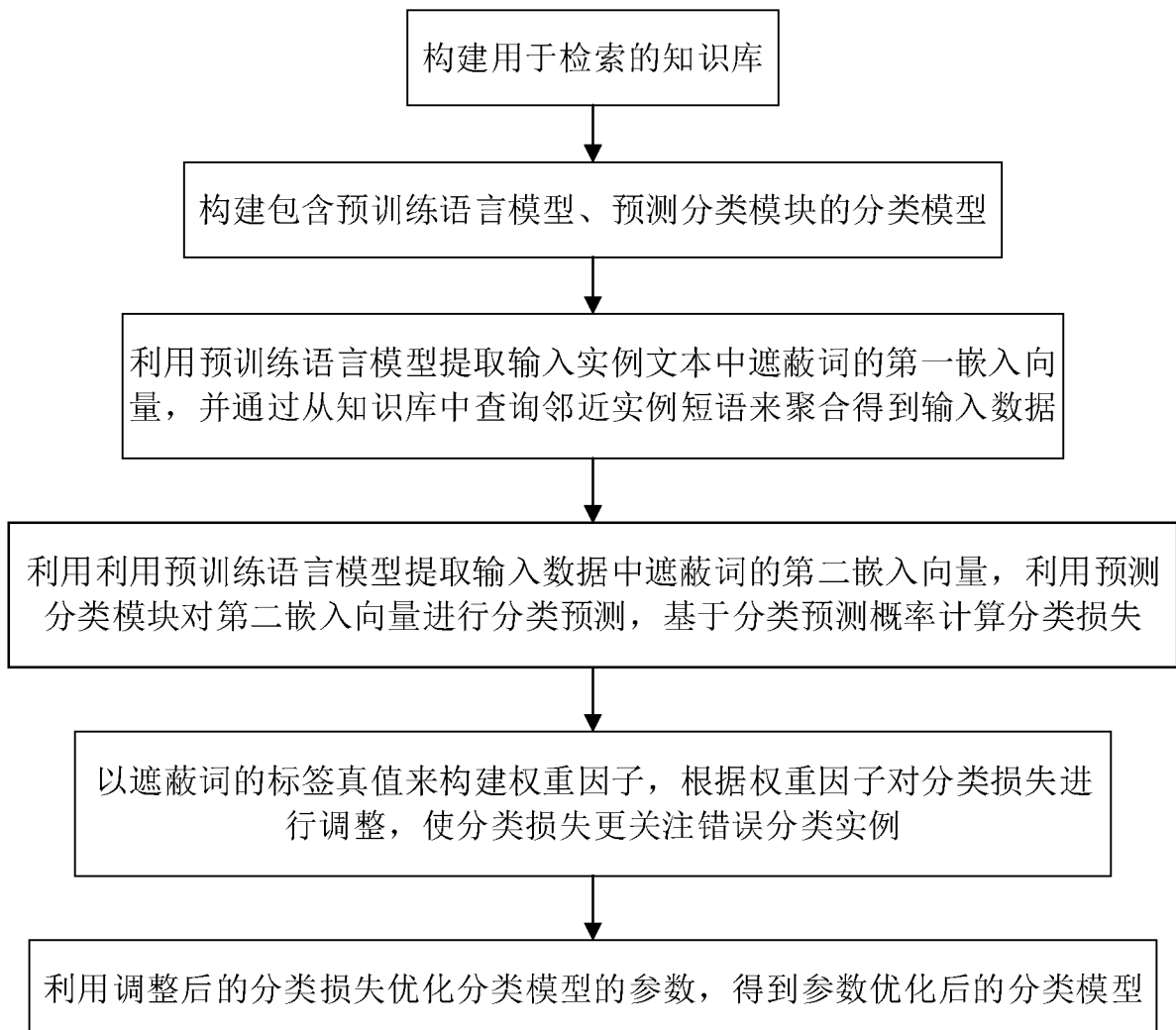
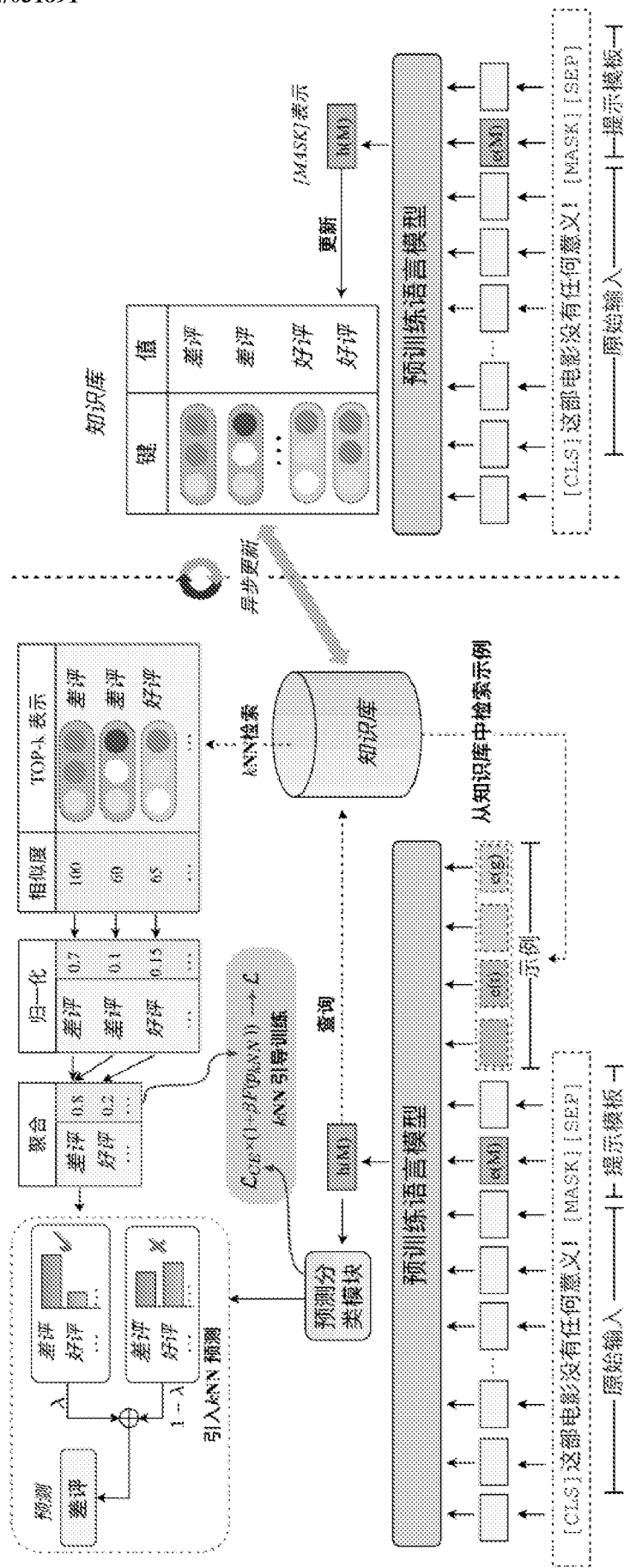


图 1



a. 检索增强的提示学习

b. 知识库的构建和更新

图 2

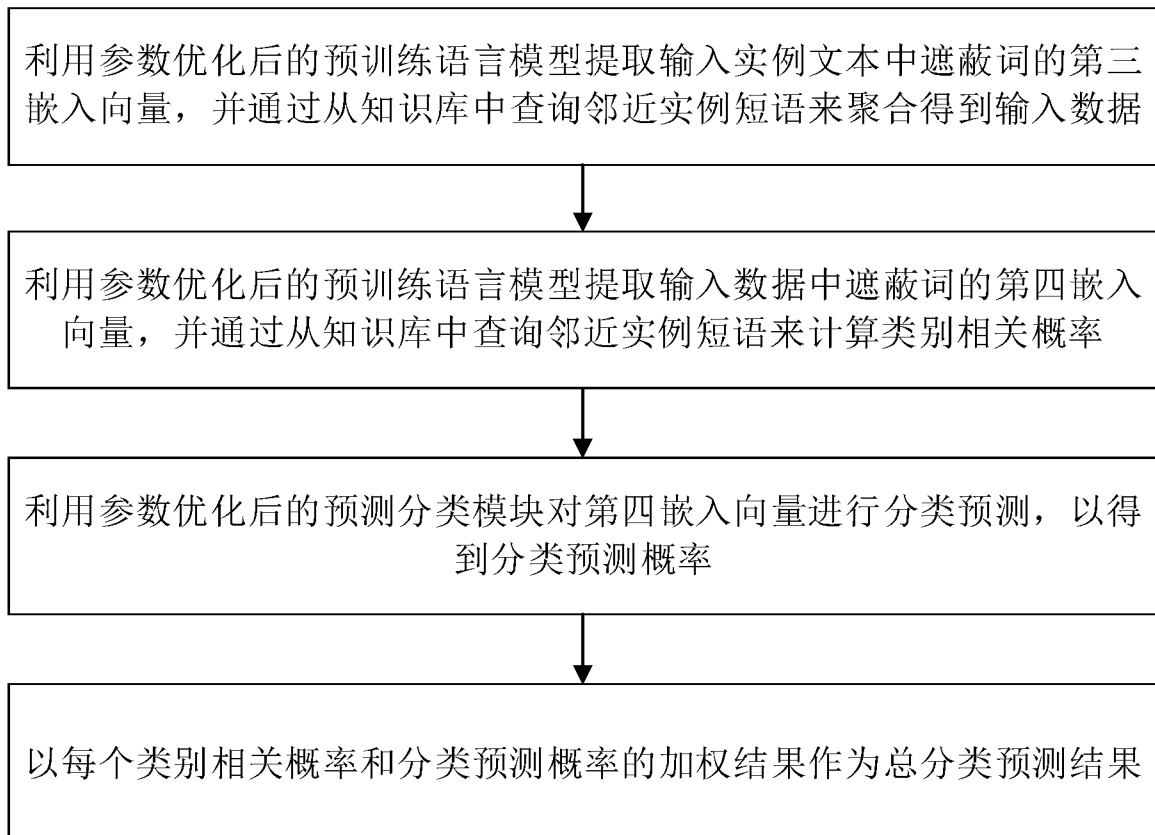


图 3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2022/137938

A. CLASSIFICATION OF SUBJECT MATTER

G06F40/30(2020.01)i;G06F16/33(2019.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS, CNTXT, EPTXT, USTXT, WOTXT, DWPI, VEN, CNKI, IEEE: 知识库, 键值, 短语, 嵌入向量, 预训练, 语言模型, 分类, 预测, 遮蔽, 掩蔽, 掩盖, 掩码, 邻近, 相邻, 搜索, 检索, 查询, 聚合, 合并, 拼接, 标签, 损失, 权重, knowledge w base, key w value, phrase, embedded w vector, train, language w model, bert, classification, predict+, mask, knn, query, search, retrieval, cluster+, combine, joint, label, tag, lose, weight

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 115270988 A (ZHEJIANG UNIVERSITY) 01 November 2022 (2022-11-01) claims 1-10	1-10
Y	CN 111401077 A (TENCENT TECHNOLOGY (SHENZHEN) CO., LTD.) 10 July 2020 (2020-07-10) description, paragraphs 0033-0207	1, 4-6
Y	WO 2022141878 A1 (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 07 July 2022 (2022-07-07) description, p. 5, paragraph 6 to p. 8, paragraph 3	1, 4-6
A	CN 112614538 A (XIAMEN UNIVERSITY) 06 April 2021 (2021-04-06) entire document	1-10
A	CN 113987209 A (ZHEJIANG UNIVERSITY) 28 January 2022 (2022-01-28) entire document	1-10
A	CN 114510572 A (FOSHAN UNIVERSITY) 17 May 2022 (2022-05-17) entire document	1-10



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“D” document cited by the applicant in the international application

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

28 February 2023

Date of mailing of the international search report

06 March 2023

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
China No. 6, Xitucheng Road, Jimenqiao, Haidian District,
Beijing 100088

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2022/137938**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 114565104 A (TENCENT TECHNOLOGY (SHENZHEN) CO., LTD. et al.) 31 May 2022 (2022-05-31) entire document	1-10
A	US 2021089724 A1 (GOOGLE L.L.C.) 25 March 2021 (2021-03-25) entire document	1-10

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2022/137938

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	115270988	A	01 November 2022	None			
CN	111401077	A	10 July 2020	HK	40025773	A0	31 December 2022
WO	2022141878	A1	07 July 2022	CN	112699216	A	23 April 2021
CN	112614538	A	06 April 2021	None			
CN	113987209	A	28 January 2022	None			
CN	114510572	A	17 May 2022	None			
CN	114565104	A	31 May 2022	None			
US	2021089724	A1	25 March 2021	DE	212020000731	U1	28 September 2022
				WO	2021061555	A1	01 April 2021
				JP	2022550326	A	01 December 2022
				CN	114514540	A	17 May 2022
				IN	202227013807	A	29 July 2022

国际检索报告

国际申请号

PCT/CN2022/137938

A. 主题的分类

G06F40/30(2020.01)i;G06F16/33(2019.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06F

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

CNABS, CNTXT, EPTXT, USTXT, WOTXT, DWPI, VEN, CNKI, IEEE: 知识库, 键值, 短语, 嵌入向量, 预训练, 语言模型, 分类, 预测, 遮蔽, 掩蔽, 掩盖, 掩码, 邻近, 相邻, 搜索, 检索, 查询, 聚合, 合并, 拼接, 标签, 损失, 权重, knowledge w base, key w value, phrase, embedded w vector, train, language w model, bert, classification, predict+, mask, knn, query, search, retriev+, cluster+, combine, joint, label, tag, lose, weight

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 115270988 A (浙江大学) 2022年11月1日 (2022 - 11 - 01) 权利要求1-10	1-10
Y	CN 111401077 A (腾讯科技(深圳)有限公司) 2020年7月10日 (2020 - 07 - 10) 说明书第0033-0207段	1, 4-6
Y	WO 2022141878 A1 (平安科技(深圳)有限公司) 2022年7月7日 (2022 - 07 - 07) 说明书第5页第6段-第8页第3段	1, 4-6
A	CN 112614538 A (厦门大学) 2021年4月6日 (2021 - 04 - 06) 全文	1-10
A	CN 113987209 A (浙江大学) 2022年1月28日 (2022 - 01 - 28) 全文	1-10
A	CN 114510572 A (佛山科学技术学院) 2022年5月17日 (2022 - 05 - 17) 全文	1-10

☒ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“D” 申请人在国际申请中引证的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2023年2月28日

国际检索报告邮寄日期

2023年3月6日

ISA/CN的名称和邮寄地址

中国国家知识产权局

中国北京市海淀区蓟门桥西土城路6号 100088

授权官员

凌燕翔

传真号 (86-10)62019451

电话号码 (+86) 010-53961323

C. 相关文件		
类 型*	引用文件，必要时，指明相关段落	相关的权利要求
A	CN 114565104 A (腾讯科技 (深圳) 有限公司等) 2022年5月31日 (2022 - 05 - 31) 全文	1-10
A	US 2021089724 A1 (GOOGLE LLC) 2021年3月25日 (2021 - 03 - 25) 全文	1-10

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2022/137938

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	115270988	A	2022年11月1日	无			
CN	111401077	A	2020年7月10日	HK	40025773	A0	2022年12月31日
WO	2022141878	A1	2022年7月7日	CN	112699216	A	2021年4月23日
CN	112614538	A	2021年4月6日	无			
CN	113987209	A	2022年1月28日	无			
CN	114510572	A	2022年5月17日	无			
CN	114565104	A	2022年5月31日	无			
US	2021089724	A1	2021年3月25日	DE	212020000731	U1	2022年9月28日
				WO	2021061555	A1	2021年4月1日
				JP	2022550326	A	2022年12月1日
				CN	114514540	A	2022年5月17日
				IN	202227013807	A	2022年7月29日