

(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 12 月 19 日 (19.12.2024)



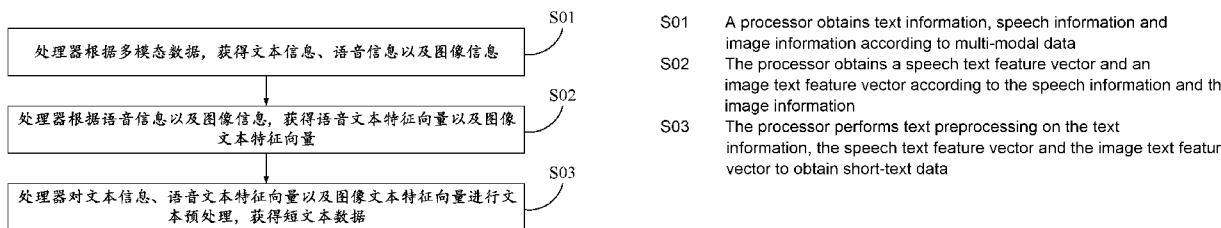
(10) 国际公布号
WO 2024/255723 A1

- (51) 国际专利分类号:
G06F 16/35 (2019.01) **G06V 10/40** (2022.01)
G10L 13/08 (2013.01)
- (21) 国际申请号: PCT/CN2024/098340
- (22) 国际申请日: 2024 年 6 月 11 日 (11.06.2024)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202310694668.X 2023年6月12日 (12.06.2023) CN
- (71) 申请人: 青岛海尔电冰箱有限公司 (QINGDAO HAIER REFRIGERATOR CO.,LTD) [CN/CN]; 中国山东省青岛市崂山区海尔路1号海尔工业园, Shandong 266101 (CN)。青岛海尔智能技术研发有限公司 (QINGDAO HAIER SMART TECHNOLOGY R&D CO., LTD.) [CN/CN]; 中国山东省青岛市崂山区海尔路1号, Shandong 266101 (CN)。海尔智家股份有限公司 (HAIER SMART HOME CO.,LTD) [CN/CN]; 中国山东省青岛市崂山区海尔路1号海尔工业园, Shandong 266101 (CN)。

(72) 发明人: 曾谁飞 (ZENG, Shuifei); 中国山东省青岛市崂山区海尔路1号海尔工业园, Shandong 266101 (CN)。刘卫强 (LIU, Weiqiang); 中国山东省青岛市崂山区海尔路1号海尔工业园, Shandong 266101 (CN)。李敏 (LI, Min); 中国山东省青岛市崂山区海尔路1号海尔工业园, Shandong 266101 (CN)。孔令磊 (KONG, Linglei); 中国山东省青岛市崂山区海尔路1号海尔工业园, Shandong 266101 (CN)。张景瑞 (ZHANG, Jingrui); 中国山东省青岛市崂山区海尔路1号海尔工业园, Shandong 266101 (CN)。谢充 (XIE, Chong); 中国山东省青岛市崂山区海尔路1号海尔工业园, Shandong 266101 (CN)。吴国章 (WU, Guozhang); 中国山东省青岛市崂山区海尔路1号海尔工业园, Shandong 266101 (CN)。窦振东 (DOU, Zhendong); 中国山东省青岛市崂山区海尔路1号海尔工业园, Shandong 266101 (CN)。谭夏霞 (TAN, Xiaxia); 中国山东省青岛市崂山区海尔路1号海尔工业园, Shandong 266101 (CN)。李京昌 (LI, Jingchang); 中国山东省青岛市崂山区海尔路1号海尔工业园, Shandong 266101 (CN)。葛宗玉 (GE, Zongyu); 中国山东省青岛市崂山区海尔路1号海尔工业园, Shandong 266101 (CN)。魏倩 (WEI, Qian); 中国山东省青岛市崂山区海尔路

(54) Title: TEXT CLASSIFICATION METHOD AND APPARATUS FOR MULTI-MODAL DATA, AND REFRIGERATION DEVICE AND MEDIUM

(54) 发明名称: 用于多模态数据的文本分类方法及装置、制冷设备、介质



(57) Abstract: The present application relates to the technical field of data processing. Disclosed is a text classification method for multi-modal data. The method comprises: obtaining text information, speech information and image information according to multi-modal data; obtaining a speech text feature vector and an image text feature vector according to the speech information and the image information; and performing text preprocessing on the text information, the speech text feature vector and the image text feature vector to obtain short-text data. Further disclosed in the present application are a text classification apparatus for multi-modal data, and a refrigeration device and a medium.

(57) 摘要: 本申请涉及数据处理技术领域, 公开一种用于多模态数据的文本分类方法, 包括: 根据多模态数据, 获得文本信息、语音信息以及图像信息; 根据语音信息以及图像信息, 获得语音文本特征向量以及图像文本特征向量; 对文本信息、语音文本特征向量以及图像文本特征向量进行文本预处理, 获得短文本数据。本申请还公开一种用于多模态数据的文本分类装置及制冷设备、介质。



1号海尔工业园, Shandong 266101 (CN)。马全亭
(MA, Quanting); 中国山东省青岛市崂山区海尔路
1号海尔工业园, Shandong 266101 (CN)。

(74) 代理人: 青岛中家标准专利代理有限公司 (QINGDAO ZHONGJIA STANDARD PATENT AGENCY CO., LTD); 中国山东省青岛市崂山区松岭路177号青岛国际创新园二期I座11层1106房间, Shandong 266000 (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, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, 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))。

用于多模态数据的文本分类方法及装置、制冷设备、介质

本申请基于申请号为 202310694668.X、申请日为 2023 年 6 月 12 日的中国专利申请提出，并要求该中国专利申请的优先权，该中国专利申请的全部内容在此引入本申请作为参考。

5 技术领域

本申请涉及数据处理技术领域，例如涉及一种用于多模态数据的文本分类方法及装置、制冷设备、介质。

背景技术

10 目前，随着多模态语音降噪技术、语音识别技术和信号处理技术的快速发展，从多模态数据中提取出用以实现交互的语音或者文本成为工业界的研究热点。在智能家居场景下，针对用户交互场景，智能认知与感知存在交互慢和反馈信息错误率高的问题，影响用户的产品体验。因此，如何从多模态数据中提取获得高质量的短文本，成为当前亟需解决的技术难题。

相关技术公开一种基于多模态深度学习的文本分类方法，包括：S01，获取实时音视频数据和历史音视频数据。S02，对实时音视频数据和历史音视频数据进行预处理，获取有效的语音数据和视频数据。S03，转写有效语音数据为语音文本数据。S04，获取有效视频数据中局部区域的视频图像，并转写视频图像为图像文本数据。S05，根据语音文本数据和图像文本数据，获得该文本数据的上下文信息和文本语义特征的权重信息。S06，将上下文信息和权重信息经全连接层组合后，输出至分类器计算得分得到分类结果信息，并
20 判断音视频数据生成文本的类型信息。S07，输出生成文本的类别信息。

在实现本公开实施例的过程中，发现相关技术中至少存在如下问题：

相关技术选取实时音视频数据和历史音视频数据作为参考数据，通过参考数据生成交互场景的文本。虽然实时音视频数据和历史音视频数据包含有文本特征，但是，多模态数据的数据来源不限于音视频数据。如此，仅采用音视频数据作为文本提取的参考数据，无法充分挖掘出多模态数据中其他数据来源的短文本特征，影响短文本提取的有效性。
25

需要说明的是，在上述背景技术部分公开的信息仅用于加强对本申请的背景的理解，因此可以包括不构成对本领域普通技术人员已知的现有技术的信息。

发明内容

30 为了对披露的实施例的一些方面有基本的理解，下面给出了简单的概括。所述概括不是泛泛评述，也不是要确定关键 / 重要组成元素或描绘这些实施例的保护范围，而是作为后面的详细说明的序言。

本公开实施例提供了一种用于多模态数据的文本分类方法、装置、制冷设备和介质，

以解决提升多模态数据短文本提取的有效性。

在一些实施例中，所述方法包括：根据多模态数据，获得文本信息、语音信息以及图像信息；根据语音信息以及图像信息，获得语音文本特征向量以及图像文本特征向量；对文本信息、语音文本特征向量以及图像文本特征向量进行文本预处理，获得短文本数据。

5 在一些实施例中，所述根据语音信息以及图像信息，获得语音文本特征向量以及图像文本特征向量，包括：对语音信息进行预处理，获得有效语音数据以及有效视频数据；对有效语音数据进行语音识别处理，获得语音文本特征向量；根据有效视频数据以及图像信息，获得图像文本特征向量。

10 在一些实施例中，所述根据有效视频数据以及图像信息，获得图像文本特征向量，包括：对有效视频数据进行帧提取处理，获得第二图像数据；对图像信息进行过滤处理，获得第一图像数据；对第一图像数据以及第二图像数据进行图像合成处理，获得图像数据；输入图像数据至第一深度神经网络模型进行模型训练，获得图像文本特征向量。

在一些实施例中，所述第一深度神经网络模型包括迁移与蒸馏扩散深度融合模型。

15 在一些实施例中，所述对有效语音数据进行语音识别处理，获得语音文本特征向量，包括：输入有效语音数据至编码器进行语音编码，获得语音编码数据；输入语音编码数据至解码器进行语音解码，获得语音文本特征向量。

20 在一些实施例中，所述对文本信息、语音文本特征向量以及图像文本特征向量进行文本预处理，获得短文本数据，包括：提取文本信息、语音文本特征向量以及图像文本特征向量各自的短文本信息；汇集文本信息、语音文本特征向量以及图像文本特征向量各自的短文本信息，获得短文本数据。

在一些实施例中，所述方法还包括：对短文本数据进行数据标注处理，获得待质检数据；在待质检数据满足质检要求的情况下，利用已训练完成的 Bert 模型对待质检数据进行文本特征提取，获得目标短文本特征数据；利用 Focal Loss 损失函数对目标短文本特征数据进行短文本分类处理，获得目标短文本数据。

25 在一些实施例中，按照以下方式训练 Bert 模型：以基准网络参数对初始 Bert 模型进行模型训练，获得模型训练后的网络参数以及排序信息；根据模型训练后的网络参数以及排序信息，确定网络参数的目标数量；利用目标数量的网络参数训练初始 Bert 模型，获得已训练完成的 Bert 模型。

30 在一些实施例中，所述利用目标数量的网络参数训练初始 Bert 模型，获得已训练完成的 Bert 模型，包括：利用目标数量的网络参数训练初始 Bert 模型，获得基准 Bert 模型；利用知识蒸馏技术将教师网络模型迁移至学生网络模型，获得基准学生网络模型。

在一些实施例中，所述利用知识蒸馏技术将教师网络模型迁移至学生网络模型，获得

基准学生网络模型之后，还包括：获得基准学生网络模型的所有编码器的层数以及各编码器的权重；根据所有编码器的层数以及各编码器的权重，确定编码器的目标层数；利用目标层数的编码器训练基准学生网络模型，获得已训练完成的 Bert 模型。

在一些实施例中，所述装置，包括：语音预处理单元，被配置为对多模态数据相关联的语音信息进行预处理，获得有效语音数据；模型训练单元，被配置为利用第一深度神经网络模型对图像数据进行模型训练，获得图像文本特征向量；语音识别单元，被配置为对有效语音数据进行语音识别处理，获得语音文本特征向量；文本预处理单元，被配置为对文本信息、语音文本特征向量以及图像文本特征向量进行文本预处理，获得短文本数据。

在一些实施例中，所述语音预处理单元还被配置为对音视频信息进行预处理，获得有效视频数据，所述装置还包括：图像帧提取单元，被配置为对有效视频数据进行帧提取处理，获得第二图像数据；图像过滤单元，被配置为对图像信息进行过滤处理，获得第一图像数据；图像融合单元，被配置为对第一图像数据以及第二图像数据进行图像合成处理，获得图像数据。

在一些实施例中，所述装置还包括：数据标注单元，被配置为对短文本数据进行数据标注处理，获得待质检数据；数据质检单元，被配置为在待质检数据满足质检要求的情况下，利用已训练完成的 Bert 模型对待质检数据进行文本特征提取，获得目标短文本特征数据；文本分类单元，被配置为利用 Focal Loss 损失函数对目标短文本特征数据进行短文本分类处理，获得目标短文本数据。

在一些实施例中，所述装置包括：处理器和存储有程序指令的存储器，所述处理器被配置为在运行所述程序指令时，执行如前述的用于多模态数据的文本分类方法。

在一些实施例中，所述制冷设备，包括：设备本体；如前述的用于多模态数据的文本分类装置，被安装于所述设备本体。

在一些实施例中，所述存储介质，存储有程序指令，所述程序指令在运行时，执行如前述的用于多模态数据的文本分类方法。

本公开实施例提供的用于多模态数据的文本分类方法、装置、制冷设备和存储介质，可以实现以下技术效果：

本公开实施例根据多模态数据获得文本信息、语音信息和图像信息。再根据语音信息和图像信息，获得语音文本特征向量和图像文本特征向量，以分别利用语音文本特征向量和图像文本特征向量体现出语音和图像各自的文本特征。最后，对文本信息、语音文本特征向量以及图像文本特征向量进行预处理获得短文本数据。如此，通过将同时携带有文本特征的语音文本特征向量和图像文本特征向量与文本信息进行预处理，本公开实施例能够从多模态数据中提取出不同数据源的短文本特征，从而获得高质量的短文本，进而提升

多模态数据短文本提取的有效性。

以上的总体描述和下文中的描述仅是示例性和解释性的，不用于限制本申请。

附图说明

一个或多个实施例通过与之对应的附图进行示例性说明，这些示例性说明和附图并不构成对实施例的限定，附图中具有相同参考数字标号的元件示为类似的元件，附图不构成比例限制，并且其中：

图 1 是本公开实施例提供的用于多模态数据的文本分类装置的示意图；

图 2 是本公开实施例提供的一个用于多模态数据的文本分类方法的示意图；

图 3 是本公开实施例提供的另一个用于多模态数据的文本分类方法的示意图；

图 4 是本公开实施例提供的另一个用于多模态数据的文本分类方法的示意图；

图 5 是本公开实施例提供的另一个用于多模态数据的文本分类方法的示意图；

图 6 是本公开实施例提供的另一个用于多模态数据的文本分类方法的示意图；

图 7 是本公开实施例的一个应用示意图；

图 8 是本公开实施例提供的一个用于多模态数据的文本分类装置的示意图；

图 9 是本公开实施例提供的另一个用于多模态数据的文本分类装置的示意图；

图 10 是本公开实施例提供的一个制冷设备的示意图。

附图标记：

10：语音预处理单元；20：图像过滤单元；30：语音识别单元；40：图像帧提取单元；

50：图像融合单元；60：模型训练单元；70：文本预处理单元；80：数据标注单元；

90：数据质检单元；100：文本分类单元；

200：用于多模态数据的文本分类装置；300：用于多模态数据的文本分类装置；

400：处理器；401：存储器；402：通信接口；403：总线；600：制冷设备。

具体实施方式

为了能够更加详尽地了解本公开实施例的特点与技术内容，下面结合附图对本公开实施例的实现进行详细阐述，所附附图仅供参考说明之用，并非用来限定本公开实施例。在以下的技术描述中，为方便解释起见，通过多个细节以提供对所披露实施例的充分理解。然而，在没有这些细节的情况下，一个或多个实施例仍然可以实施。在其它情况下，为简化附图，熟知的结构和装置可以简化展示。

本公开实施例的说明书和权利要求书及上述附图中的术语“第一”、“第二”等是用于区别类似的对象，而不必用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换，以便这里描述的本公开实施例的实施例。此外，术语“包括”和“具有”以及他们的任何变形，意图在于覆盖不排他的包含。

除非另有说明，术语“多个”表示两个或两个以上。

本公开实施例中，字符“/”表示前后对象是一种“或”的关系。例如，A/B 表示：A 或 B。

术语“和/或”是一种描述对象的关联关系，表示可以存在三种关系。例如，A 和/或 B，表示：A 或 B，或，A 和 B 这三种关系。

5 术语“对应”可以指的是一种关联关系或绑定关系，A 与 B 相对应指的是 A 与 B 之间是一种关联关系或绑定关系。

结合图 1 和图 8 所示，本公开实施例提供一种用于多模态数据的文本分类装置 200，包括语音预处理单元 10、模型训练单元 60、语音识别单元 30 以及文本预处理单元 70。语音预处理单元 10，被配置为对多模态数据相关联的语音信息进行预处理，获得有效语音数据。模型训练单元 60，被配置为利用第一深度神经网络模型对图像数据进行模型训练，获得图像文本特征向量。语音识别单元 30，被配置为对有效语音数据进行语音识别处理，获得语音文本特征向量。文本预处理单元 70，被配置为对文本信息、语音文本特征向量以及图像文本特征向量进行文本预处理，获得短文本数据。

15 可选地，语音预处理单元 10 还被配置为对音视频信息进行预处理，获得有效视频数据。用于多模态数据的文本分类系统还包括图像帧提取单元 40、图像过滤单元 20 以及图像融合单元 50。图像帧提取单元 40，被配置为对有效视频数据进行帧提取处理，获得第二图像数据。图像过滤单元 20，被配置为对图像信息进行过滤处理，获得第一图像数据。图像融合单元 50，被配置为对第一图像数据以及第二图像数据进行图像合成处理，获得图像数据。

20 可选地，用于多模态数据的文本分类系统还包括数据标注单元 80、数据质检单元 90 以及文本分类单元 100。数据标注单元 80，被配置为对短文本数据进行数据标注处理，获得待质检数据。数据质检单元 90，被配置为在待质检数据满足质检要求的情况下，利用已训练完成的 Bert 模型对待质检数据进行文本特征提取，获得目标短文本特征数据。文本分类单元 100，被配置为利用 Focal Loss 损失函数对目标短文本特征数据进行短文本分类处理，获得目标短文本数据。

25 基于上述用于多模态数据的文本分类系统，结合图 2 所示，本公开实施例提供一种用于多模态数据的文本分类方法，包括：

S01，处理器根据多模态数据，获得文本信息、语音信息以及图像信息。

该步骤中，多模态数据的类型包括文本信息、语音信息以及图像信息中的部分或者全部。处理器按照以下方式获得多模态数据：处理器通过人机交互模式获取多模态数据。人机交互模式包括但不限于蓝牙、Wi-Fi（Wireless Fidelity，无线保真）、Internet（因特网）、Web（World Wide Web，万维网）、麦克风阵列、拾音器以及软件。其中，软件包括

APP(Application, 应用程序)、信息采集工具以及触屏中的一个或者多个。

S02, 处理器根据语音信息以及图像信息, 获得语音文本特征向量以及图像文本特征向量。

S03, 处理器对文本信息、语音文本特征向量以及图像文本特征向量进行文本预处理, 获得短文本数据。

采用本公开实施例提供的用于多模态数据的文本分类方法, 本公开实施例根据多模态数据获得文本信息、语音信息和图像信息。再根据语音信息和图像信息, 获得语音文本特征向量和图像文本特征向量, 以分别利用语音文本特征向量和图像文本特征向量体现出语音和图像各自的文本特征。最后, 对文本信息、语音文本特征向量以及图像文本特征向量进行预处理获得短文本数据。如此, 通过将同时携带有文本特征的语音文本特征向量和图像文本特征向量与文本信息进行预处理, 能够从多模态数据中提取出不同数据源的短文本特征, 从而获得高质量的短文本, 进而提升多模态数据短文本提取的有效性。

可选地, 处理器按照以下方式从多模态数据中提取获得文本信息: 处理器从多模态数据中获得初始文本数据。处理器对初始文本数据进行文本预处理, 获得文本信息。其中, 文本预处理包括分词、词形归一化以及删除停用词。

可选地, 结合图 3 所示, 处理器根据语音信息以及图像信息, 获得语音文本特征向量以及图像文本特征向量, 包括:

S11, 处理器对语音信息进行预处理, 获得有效语音数据以及有效视频数据。

该步骤中, 处理器对语音信息进行预处理, 获得有效语音数据以及有效视频数据, 包括:

处理器对多模态数据依次进行数据清洗以及格式转换处理, 提取有效声音数据。处理器对有效声音数据进行语音视频分离处理, 获得语音预处理数据以及有效视频数据。处理器对语音预处理数据进行有效时长处理, 获得有效语音数据。如此, 通过进行数据清洗以及格式转换处理, 能够去除多模态数据中的部分干扰并统一数据格式, 获得有效声音数据。通过对有效声音数据进行语音视频分离处理可将语音预处理数据和有效视频数据进行分离, 获得语音预处理数据。通过对语音预处理数据进行有效时长处理, 能够有效处理或者消除空语音以及背景音等。其中, 背景音包括机器运行阶段的机器声音。在实际应用中, 处理器可通过数据处理工具或者数据处理脚本对多模态数据进行数据清洗、格式转换处理以及语音视频分离处理。

S12, 处理器对有效语音数据进行语音识别处理, 获得语音文本特征向量。

该步骤中, 处理器对有效语音数据进行语音识别处理, 获得语音文本特征向量, 包括: 处理器输入有效语音数据至第一编码器进行编码处理, 获得语音编码数据。处理器利用第

二深度神经网络模型对语音编码数据进行特征提取，获得语音文本特征向量。可选地，第一编码器为 Wav2VEC 编码器。第二深度神经网络配置有多头注意力机制。如此，通过采用配置多头注意力机制的第二深度神经网络，实现语音编码数据的向量空间的映射，并，提取出语音所涵盖的语义、语法和语境特征，并通过语音文本特征向量予以体现

5 S13，处理器根据有效视频数据以及图像信息，获得图像文本特征向量。

这样，本公开实施例通过对语音信息进行预处理，可从语音信息中提取出有效语音数据和有效视频数据，降低多模态数据中的干扰并实现数据格式的统一。此外，可有效地消除空语音和背景音。提升有效语音数据和有效视频数据识别的有效性。在此基础上，本公开实施例再对有效语音数据进行语音识别处理，获得语音文本特征向量。有利于从语音中
10 充分挖掘语音文本特征。并根据有效视频数据以及图像信息，获得图像文本特征向量，有利于从图像中充分挖掘图像文本特征。

可选地，结合图 4 所示，处理器根据有效视频数据以及图像信息，获得图像文本特征向量，包括：

S21，处理器对有效视频数据进行帧提取处理，获得第二图像数据。

15 S22，处理器对图像信息进行过滤处理，获得第一图像数据。

S23，处理器对第一图像数据以及第二图像数据进行图像合成处理，获得图像数据。

S24，处理器输入图像数据至第一深度神经网络模型进行模型训练，获得图像文本特征向量。

这样，本公开实施例先通过对有效视频数据进行帧提取以从视频数据中提取出第二图像数据，并对图像信息进行过滤处理以从图像信息中提取出第一图像数据。再对第一图像数据和第二图像数据进行图像合成处理以将第一图像数据和第二图像进行有效的融合，生成图像数据。最后，将融合获得的图像数据输入至第一深度神经网络模型进行模型训练，有利于从图像数据中准确地识别语义特征。

可选地，第一深度神经网络模型包括迁移与蒸馏扩散深度融合模型。

25 其中，迁移与蒸馏扩散深度融合模型表示利用迁移学习，采用预先训练好的复杂模型的输出作为监督信号去训练一简单的网络模型。通常情况下，复杂模型被称为教师模型。简单的网络模型被称为学生模型。

这样，通过选用迁移与蒸馏扩散深度融合模型作为第一深度神经网络模型，可利用该模型的时空以及长距离依赖特性，将教师模型引入时间与空间连续的学生模型，更有利于
30 从图像数据中挖掘出丰富的语义特征，进一步提升图像文本特征向量的有效性。

可选地，处理器对有效语音数据进行语音识别处理，获得语音文本特征向量，包括：输入有效语音数据至编码器进行语音编码，获得语音编码数据。

输入语音编码数据至解码器进行语音解码，获得语音文本特征向量。

这样，本公开实施例通过输入有效语音数据至编码器进行语音编码，使得编码所获得的语音编码数据不会损失有用语音数据，有利于后续的语音文本特征的提取。

可选地，处理器输入有效语音数据至编码器进行语音编码，获得语音编码数据，包括：

- 5 处理器利用 Wav2VEC 编码器对有效语音数据进行编码处理，获得语音编码数据。这样，通过采用 Wav2VEC 编码器对有效语音数据进行编码，能够增强冗余语音数据的消除效果，从而获得更为纯净的语音编码数据。

可选地，处理器输入语音编码数据至解码器进行语音解码，获得语音文本特征向量，包括：处理器输入语音编码信号至 Transformer 网络模型进行解码处理，获得语音文本特征向量。这样，利用 Transformer 网络模型对语音编码信号进行解码处理，能够进一步增强冗余语音数据的消除效果，从而充分提取出语音文本特征。

可选地，处理器对文本信息、语音文本特征向量以及图像文本特征向量进行文本预处理，获得短文本数据，包括：

处理器提取文本信息、语音文本特征向量以及图像文本特征向量各自的短文本信息。

- 15 处理器汇集文本信息、语音文本特征向量以及图像文本特征向量各自的短文本信息，获得短文本数据。

这样，本公开实施例通过从文本信息、语音文本特征向量和图像文本特征向量中分别提取出各自的短文本信息，再汇集上述不同数据来源的短文本信息。实现不同数据来源的短文本特征的有效提取，有利于获得高质量的短文本，从而提升多模态数据短文本提取的有效性。

结合图 5 所示，本公开实施例还提供一种用于多模态数据的文本分类方法，包括：

S31，处理器根据多模态数据，获得文本信息、语音信息以及图像信息。

S32，处理器根据语音信息以及图像信息，获得语音文本特征向量以及图像文本特征向量。

- 25 S33，处理器对文本信息、语音文本特征向量以及图像文本特征向量进行文本预处理，获得短文本数据。

S34，处理器对短文本数据进行数据标注处理，获得待质检数据。

该步骤中，处理器对短文本数据进行数据标注处理，获得待质检数据，包括：处理器对短文本数据标注槽位以及元数据属性标签，获得待质检数据。如此，为后续的数据质检提供依据。

30 S35，处理器在待质检数据满足质检要求的情况下，利用已训练完成的 Bert 模型对待质检数据进行文本特征提取，获得目标短文本特征数据。

该步骤中，处理器按照以下方式确定待之间数据满足质检要求：元数据属性标签以及食材属性标签正确且错误率小于或者等于 2%。

S36，处理器利用 Focal Loss 损失函数对目标短文本特征进行短文本分类处理，获得目标短文本数据。

5 采用本公开实施例提供的用于多模态数据的文本分类方法，本公开实施例在获得短文本数据后，依次进行数据标注以及数据质检，并在满足质检要求时，利用已训练完成的 Bert 模型对待质检数据进行文本特征提取，获得目标短文本特征数据，从而有利于从待质检数据中提取出包含丰富短文本特征的短文本特征数据。由于短文本特征数据存在类型不均衡的问题，影响短文本分类处理的分类准确率。为此，本公开实施例利用 Focal Loss 损失函数对目标短文本特征进行短文本分类处理，以降低因短文本特征数据的类型不均衡而对短文本分类处理的准确率产生的影响，进一步保证短文本的质量。

10 可选地，处理器利用 Focal Loss 损失函数对目标短文本特征进行短文本分类处理，获得目标短文本数据，包括：处理器利用 Focal Loss 损失函数对目标短文本特征数据进行短文本分类处理，获得短文本分类数据。处理器对短文本分类数据进行数据增强处理，获得目标短文本数据。这样，通过对短文本分类数据进行数据增强处理，能够扩充有限的短文本分类数据集，提升 Bert 模型的训练效果，增强模型的泛化能力。

可选地，结合图 6 所示，处理器按照以下方式训练 Bert 模型：

S41，处理器以基准网络参数对初始 Bert 模型进行模型训练，获得模型训练后的网络参数以及排序信息。

20 S42，处理器根据模型训练后的网络参数以及排序信息，确定网络参数的目标数量。

该步骤中，处理器根据模型训练后的网络参数以及排序信息，确定网络参数的目标数量，包括：处理器根据排序信息，选取排序前 α 的网络参数的数量为目标数量。处理器计算 $N=\alpha\times N_0$ 。 $0.5\leq\alpha\leq 0.8$ 。其中， N_0 表示网络参数的总数量。 N 表示网络参数的目标数量。 α 表示第一权重值。优选的， $\alpha=0.75$ 。

25 S43，处理器利用目标数量的网络参数训练初始 Bert 模型，获得已训练完成的 Bert 模型。

这样，针对初始 Bert 模型进行网络参数的数量的截取，实现对初始 Bert 模型的微调，保留了模型权重高的网络层数与网络参数，能够减小模型对内存空间的占用。

可选地，处理器按照以下方式获得基准网络参数：

30 处理器从目标短文本数据中提取获得训练数据。

处理器利用训练数据对初始 Bert 模型进行模型训练，获得基准网络参数。

这样，有利于提升基准网络参数的准确性，提升 Bert 模型的训练效果。

可选地，处理器利用目标数量的网络参数训练初始 Bert 模型，获得已训练完成的 Bert 模型，包括：

处理器利用目标数量的网络参数训练初始 Bert 模型，获得基准 Bert 模型。处理器利用知识蒸馏技术将教师网络模型迁移至学生网络模型，获得基准学生网络模型。其中，基准 Bert 模型为教师网络模型。基准学生网络模型为已训练完成的 Bert 模型。如此，通过教师网络模型至学生网络模型的迁移，以利用大模型指导小模型进行模型训练，所获得的基准学生网络模型具有与教师网络模型相当的性能，同时，大幅降低网络参数的数量，实现模型的压缩与加速。

可选地，处理器利用知识蒸馏技术将教师网络模型迁移至学生网络模型，获得基准学生网络模型之后，还包括：处理器获得基准学生网络模型的所有编码器的层数以及各编码器的权重。处理器根据所有编码器的层数以及各编码器的权重，确定编码器的目标层数。处理器利用目标层数的编码器训练基准学生网络模型，获得已训练完成的 Bert 模型。

这样，经过裁剪后的 Bert 模型的编码器层数有所降低，能减小模型对内存空间的占用，提高 Bert 模型的模型训练的效率。

可选地，处理器根据所有编码器的层数以及各编码器的权重，确定编码器的目标层数，包括：处理器按照权重大小对所有编码器的权重进行降序排列。处理器从降序排列的权重中，选取排序前 β 位的编码器的层数为目标层数。处理器计算 $M = \beta \times M_0$ 。 $0.5 \leq \beta \leq 0.8$ 。其中， M_0 表示所有编码器的层数。 M 表示编码器的目标层数。 β 表示第二权重值。优选的， $\beta = 0.75$ 。

在实际应用中，如图 7 所示，用于多模态数据的文本分类方法包括以下步骤：

S101：处理器采集多模态数据。

S102：处理器对多模态数据进行数据清洗，获得语音信息、文本信息和图像信息。

S103：处理器对语音信息进行进行预处理，获得有效语音数据以及有效视频数据。

S1041：处理器输入有效语音数据依次经 Wav2 编码器、Transformer 解码器，获得语音文本特征向量。

S1042：处理器对有效视频数据进行图像帧提取，获得第二目标图像。

S1043：处理器对图像信息进行图像过滤，获得第一目标图像。

S1044：处理器对第二目标图像和第一目标图像进行图像合成，获得图像数据。

S1045：处理器输入图像数据至迁移与蒸馏扩散深度融合模型，进行向量空间的映射处理，以从图像数据中提取出来图像文本特征向量。

S105：处理器对文本信息、语音文本特征向量以及图像文本特征向量进行文本预处理，获得短文本数据。

S106：处理器对短文本数据进行数据标注处理，获得待质检数据。

S107: 处理器判断待质检数据是否满足质检要求, 若是, 则执行 S108, 否则, 执行 S109。

S108: 处理器将待质检数据作为文本分类结果进行输出。

5 S109: 处理器利用已训练完成的 Bert 模型对待质检数据进行文本特征提取, 获得目标短文本特征数据。

S110: 处理器利用 Focal Loss 损失函数对目标短文本特征进行短文本分类处理, 获得目标短文本数据, 并将目标短文本数据作为文本分类结果进行输出。

结合图 9 所示, 本公开实施例提供一种用于多模态数据的文本分类装置 300, 包括处理器 (processor) 400 和存储器 (memory) 401。可选地, 该装置还可以包括通信接口
10 (Communication Interface) 402 和总线 403。其中, 处理器 400、通信接口 402、存储器 401 可以通过总线 403 完成相互间的通信。通信接口 402 可以用于信息传输。处理器 400 可以调用存储器 401 中的逻辑指令, 以执行上述实施例的用于多模态数据的文本分类方法。

此外, 上述的存储器 401 中的逻辑指令可以通过软件功能单元的形式实现并作为独立的产品销售或使用, 可以存储在一个计算机可读取存储介质中。

15 存储器 401 作为一种计算机可读存储介质, 可用于存储软件程序、计算机可执行程序, 如本公开实施例中的方法对应的程序指令/模块。处理器 400 通过运行存储在存储器 401 中的程序指令/模块, 从而执行功能应用以及数据处理, 即实现上述实施例中用于多模态数据的文本分类方法。

存储器 401 可包括存储程序区和存储数据区, 其中, 存储程序区可存储操作系统、至少一个功能所需的应用程序; 存储数据区可存储根据终端设备的使用所创建的数据等。此外, 存储器 401 可以包括高速随机存取存储器, 还可以包括非易失性存储器。

结合图 10 所示, 本公开实施例提供了一种制冷设备 600, 包括: 设备本体, 以及上述的用于多模态数据的文本分类装置 200(300)。用于多模态数据的文本分类装置 200(300) 被安装于设备本体。这里所表述的安装关系, 并不仅限于在产品内部放置, 还包括了与产
25 品的其他元器件的安装连接, 包括但不限于物理连接、电性连接或者信号传输连接等。本领域技术人员可以理解的是, 用于多模态数据的文本分类装置 200(300) 可以适配于可行的产品主体, 进而实现其他可行的实施例。

本公开实施例提供了一种计算机可读存储介质, 存储有计算机可执行指令, 所述计算机可执行指令设置为执行上述用于多模态数据的文本分类方法。

30 本公开实施例提供了一种计算机程序, 当所述计算机程序被计算机执行时, 使所述计算机实现上述用于多模态数据的文本分类方法。

本公开实施例提供了一种计算机程序产品, 所述计算机程序产品包括存储在计算机可

读存储介质上的计算机指令，当所述程序指令被计算机执行时，使所述计算机实现上述用于多模态数据的文本分类方法。

上述的计算机可读存储介质可以是暂态计算机可读存储介质，也可以是非暂态计算机可读存储介质。

5 本公开实施例的技术方案可以以软件产品的形式体现出来，该计算机软件产品存储在一个存储介质中，包括一个或多个指令用以使得一台计算机设备（可以是个人计算机，服务器，或者网络设备）执行本公开实施例所述方法的全部或部分步骤。而前述的存储介质可以是非暂态存储介质，包括：U 盘、移动硬盘、只读存储器（ROM，Read-Only Memory）、随机存取存储器（RAM，Random Access Memory）、磁碟或者光盘等多种可以存储程序代码的介质，也可以是暂态存储介质。

说明书中对任何现有技术的引用不是也不应被视为承认或以任何形式暗示该现有技术构成申请地区或任何其他司法管辖区的公知常识的一部分，或者该现有技术可以被本领域技术人员合理地理解和视为相关。

15 以上描述和附图充分地示出了本公开的实施例，以使本领域的技术人员能够实践它们。其他实施例可以包括结构的、逻辑的、电气的、过程的以及其他的改变。实施例仅代表可能的变化。除非明确要求，否则单独的部件和功能是可选的，并且操作的顺序可以变化。一些实施例的部分和特征可以被包括在或替换其他实施例的部分和特征。而且，本申请中使用的用词仅用于描述实施例并且不用于限制权利要求。如在实施例以及权利要求的描述中使用的，除非上下文清楚地表明，否则单数形式的“一个”（a）、“一个”（an）和“所述”（the）旨在同样包括复数形式。类似地，如在本申请中所使用的术语“和/或”是指包含一个或一个以上相关联的列出的任何以及所有可能的组合。另外，当用于本申请中时，术语“包括”（comprise）及其变型“包括”（comprises）和/或包括（comprising）等指陈述的特征、整体、步骤、操作、元素，和/或组件的存在，但不排除一个或一个以上其它特征、整体、步骤、操作、元素、组件和/或这些的分组的存在或添加。在没有更多限制的情况下，由语句“包括一个...”限定的要素，并不排除在包括所述要素的过程、方法或者设备中还存在另外的相同要素。本文中，每个实施例重点说明的可以是与其他实施例的不同之处，各个实施例之间相同相似部分可以互相参见。对于实施例公开的方法、产品等而言，如果其与实施例公开的方法部分相对应，那么相关之处可以参见方法部分的描述。

30 本领域技术人员可以意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，能够以电子硬件、或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行，可以取决于技术方案的特定应用和设计约束条件。所述技术人员可以对每个特定的应用来使用不同方法以实现所描述的功能，但是这种实现不应认为超出本

公开实施例的范围。所述技术人员可以清楚地了解到，为描述的方便和简洁，上述描述的系统、装置和单元的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

本文所披露的实施例中，所揭露的方法、产品（包括但不限于装置、设备等），可以通过其它的方式实现。例如，以上所描述的装置实施例仅仅是示意性的，例如，所述单元的划分，可以仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个单元或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另外，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口，装置或单元的间接耦合或通信连接，可以是电性，机械或其它的形式。所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例。另外，在本公开实施例中的各功能单元可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个单元中。

附图中的流程图和框图显示了根据本公开实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上，流程图或框图中的每个方框可以代表一个模块、程序段或代码的一部分，所述模块、程序段或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。在有些作为替换的实现中，方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如，两个连续的方框实际上可以基本并行地执行，它们有时也可以按相反的顺序执行，这可以依所涉及的功能而定。在附图中的流程图和框图所对应的描述中，不同的方框所对应的操作或步骤也可以以不同于描述中所披露的顺序发生，有时不同的操作或步骤之间不存在特定的顺序。例如，两个连续的操作或步骤实际上可以基本并行地执行，它们有时也可以按相反的顺序执行，这可以依所涉及的功能而定。框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合，可以用执行规定的功能或动作的专用的基于硬件的系统来实现，或者可以用专用硬件与计算机指令的组合来实现。

权利要求

1、一种用于多模态数据的文本分类方法，其特征在于，包括：

根据多模态数据，获得文本信息、语音信息以及图像信息；

根据语音信息以及图像信息，获得语音文本特征向量以及图像文本特征向量；

5 对文本信息、语音文本特征向量以及图像文本特征向量进行文本预处理，获得短文本数据。

2、根据权利要求 1 所述的方法，其中，所述根据语音信息以及图像信息，获得语音文本特征向量以及图像文本特征向量，包括：

对语音信息进行预处理，获得有效语音数据以及有效视频数据；

10 对有效语音数据进行语音识别处理，获得语音文本特征向量；

根据有效视频数据以及图像信息，获得图像文本特征向量。

3、根据权利要求 2 所述的方法，其中，所述根据有效视频数据以及图像信息，获得图像文本特征向量，包括：

对有效视频数据进行帧提取处理，获得第二图像数据；

15 对图像信息进行过滤处理，获得第一图像数据；

对第一图像数据以及第二图像数据进行图像合成处理，获得图像数据；

输入图像数据至第一深度神经网络模型进行模型训练，获得图像文本特征向量。

4、根据权利要求 3 所述的方法，其中，所述第一深度神经网络模型包括迁移与蒸馏扩散深度融合模型。

20 5、根据权利要求 2 所述的方法，其中，所述对有效语音数据进行语音识别处理，获得语音文本特征向量，包括：

输入有效语音数据至编码器进行语音编码，获得语音编码数据；

输入语音编码数据至解码器进行语音解码，获得语音文本特征向量。

25 6、根据权利要求 1 至 5 任一项所述的方法，其中，所述对文本信息、语音文本特征向量以及图像文本特征向量进行文本预处理，获得短文本数据，包括：

提取文本信息、语音文本特征向量以及图像文本特征向量各自的短文本信息；

汇集文本信息、语音文本特征向量以及图像文本特征向量各自的短文本信息，获得短文本数据。

7、根据权利要求 1 至 6 任一项所述的方法，其中，还包括：

30 对短文本数据进行数据标注处理，获得待质检数据；

在待质检数据满足质检要求的情况下，利用已训练完成的 Bert 模型对待质检数据

进行文本特征提取，获得目标短文本特征数据；

利用 Focal Loss 损失函数对目标短文本特征数据进行短文本分类处理，获得目标短文本数据。

8、根据权利要求 7 所述的方法，其中，按照以下方式训练 Bert 模型：

5 以基准网络参数对初始 Bert 模型进行模型训练，获得模型训练后的网络参数以及排序信息；

根据模型训练后的网络参数以及排序信息，确定网络参数的目标数量；

利用目标数量的网络参数训练初始 Bert 模型，获得已训练完成的 Bert 模型。

9、根据权利要求 8 所述的方法，其中，所述利用目标数量的网络参数训练初始
10 Bert 模型，获得已训练完成的 Bert 模型，包括：

利用目标数量的网络参数训练初始 Bert 模型，获得基准 Bert 模型；

利用知识蒸馏技术将教师网络模型迁移至学生网络模型，获得基准学生网络模型。

10、根据权利要求 9 所述的方法，其中，所述利用知识蒸馏技术将教师网络模型
迁移至学生网络模型，获得基准学生网络模型之后，还包括：

15 获得基准学生网络模型的所有编码器的层数以及各编码器的权重；

根据所有编码器的层数以及各编码器的权重，确定编码器的目标层数；

利用目标层数的编码器训练基准学生网络模型，获得已训练完成的 Bert 模型。

11、一种用于多模态数据的文本分类装置，其特征在于，包括：

20 语音预处理单元，被配置为对多模态数据相关联的语音信息进行预处理，获得有效语音数据；

模型训练单元，被配置为利用第一深度神经网络模型对图像数据进行模型训练，
获得图像文本特征向量；

语音识别单元，被配置为对有效语音数据进行语音识别处理，获得语音文本特征
向量；

25 文本预处理单元，被配置为对文本信息、语音文本特征向量以及图像文本特征向
量进行文本预处理，获得短文本数据。

12、根据权利要求 11 所述的装置，其中，所述语音预处理单元还被配置为对音视
频信息进行预处理，获得有效视频数据，所述装置还包括：

图像帧提取单元，被配置为对有效视频数据进行帧提取处理，获得第二图像数据；

30 图像过滤单元，被配置为对图像信息进行过滤处理，获得第一图像数据；

图像融合单元，被配置为对第一图像数据以及第二图像数据进行图像合成处理，

获得图像数据。

13、根据权利要求 11 所述的装置，其中，所述装置还包括：

数据标注单元，被配置为对短文本数据进行数据标注处理，获得待质检数据；

5 数据质检单元，被配置为在待质检数据满足质检要求的情况下，利用已训练完成的 Bert 模型对待质检数据进行文本特征提取，获得目标短文本特征数据；

文本分类单元，被配置为利用 Focal Loss 损失函数对目标短文本特征数据进行短文本分类处理，获得目标短文本数据。

14、一种用于多模态数据的文本分类装置，包括处理器和存储有程序指令的存储器，其特征在于，所述处理器被配置为在运行所述程序指令时，执行如权利要求 1 至
10 10 任一项所述的用于多模态数据的文本分类方法。

15、一种制冷设备，其特征在于，包括：

设备本体；

如权利要求 11 至 14 任一项所述的用于多模态数据文本分类装置，被安装于所述设备本体。

15 16、一种存储介质，存储有程序指令，其特征在于，所述程序指令在运行时，执行如权利要求 1 至 10 任一项所述的用于多模态数据的文本分类方法。

17、一种计算机程序，当所述计算机程序被计算机执行时，使所述计算机实现如权利要求 1 至 10 任一项所述的用于多模态数据的文本分类方法。

20 18、一种计算机程序产品，所述计算机程序产品包括存储在计算机可读存储介质上的计算机指令，当所述程序指令被计算机执行时，使所述计算机实现如权利要求 1 至 10 任一项所述的用于多模态数据的文本分类方法。

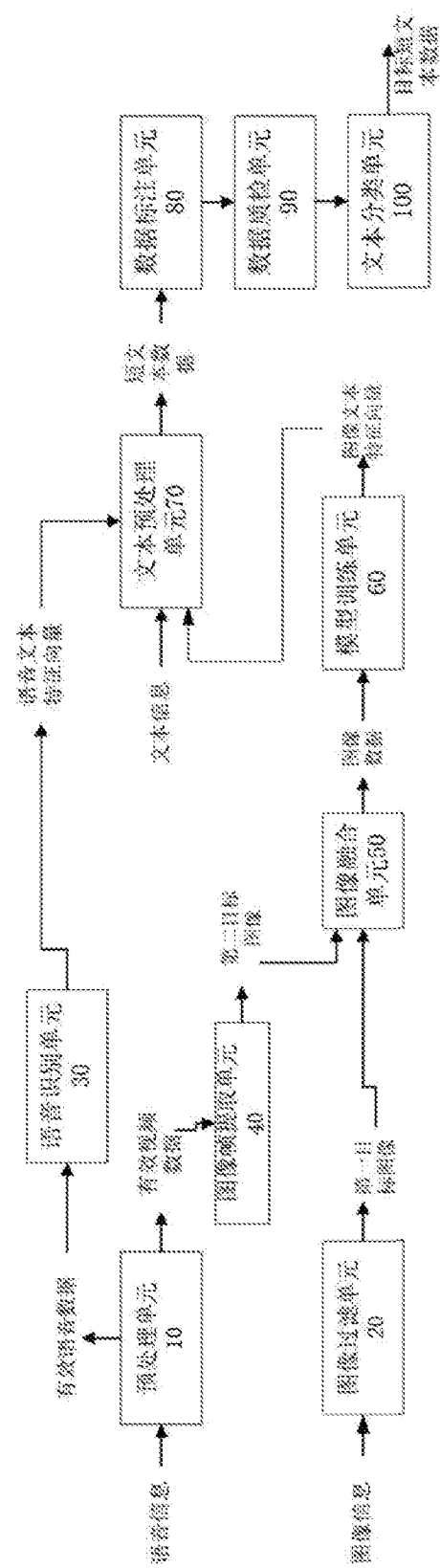
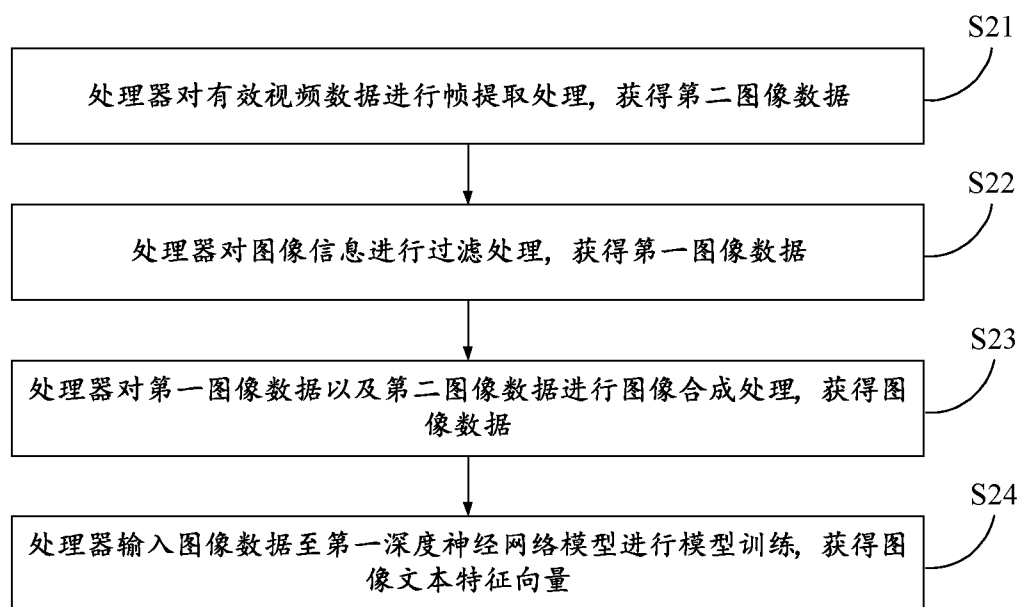
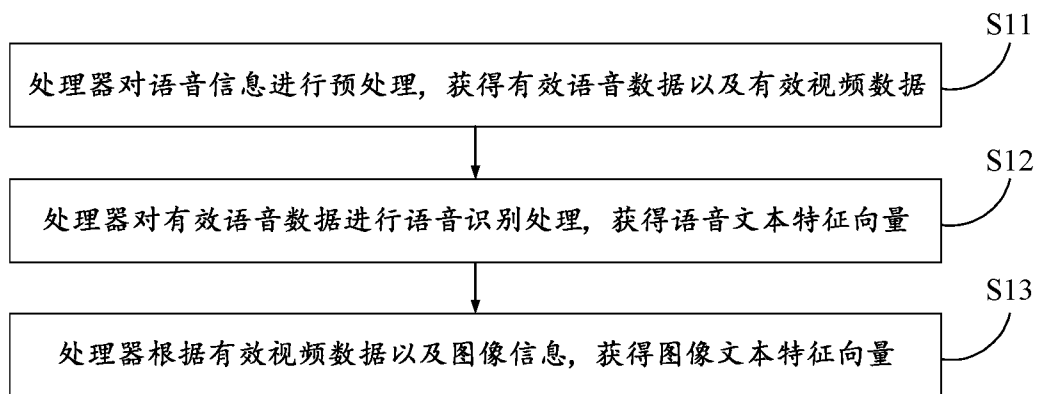
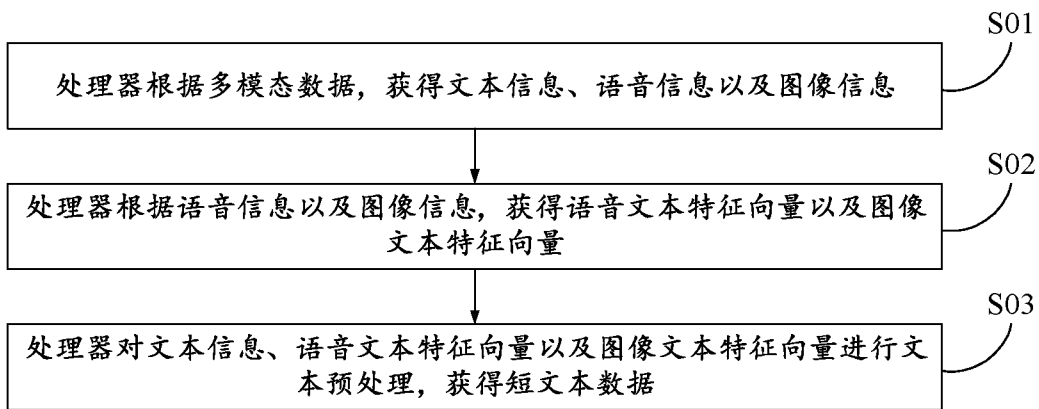


图 1



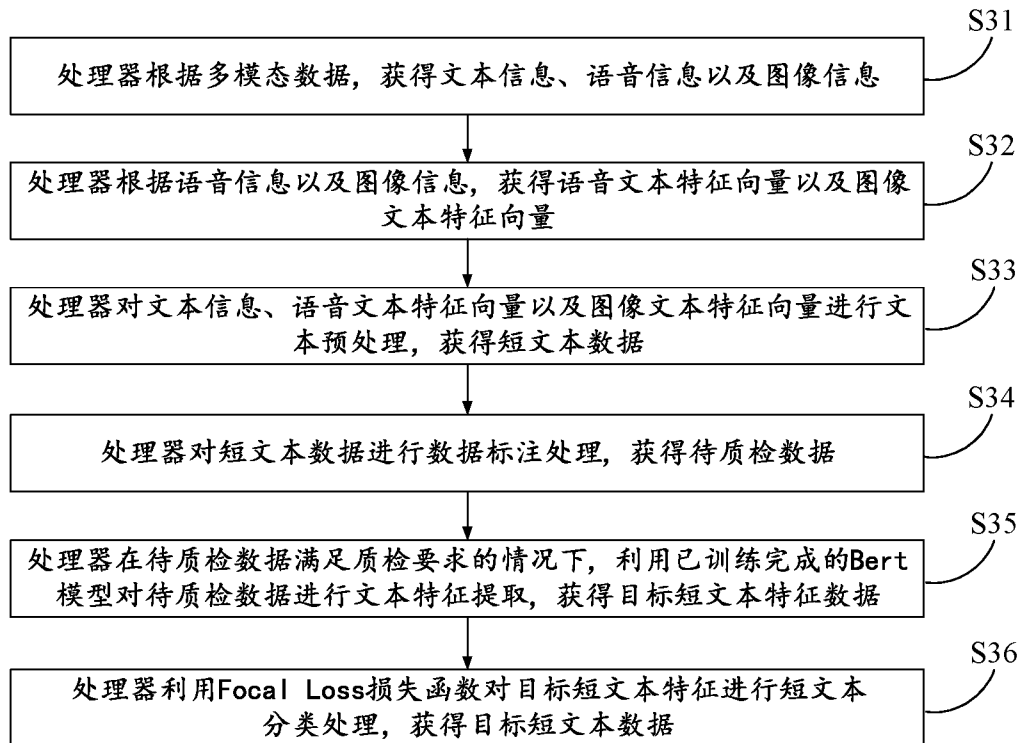


图 5

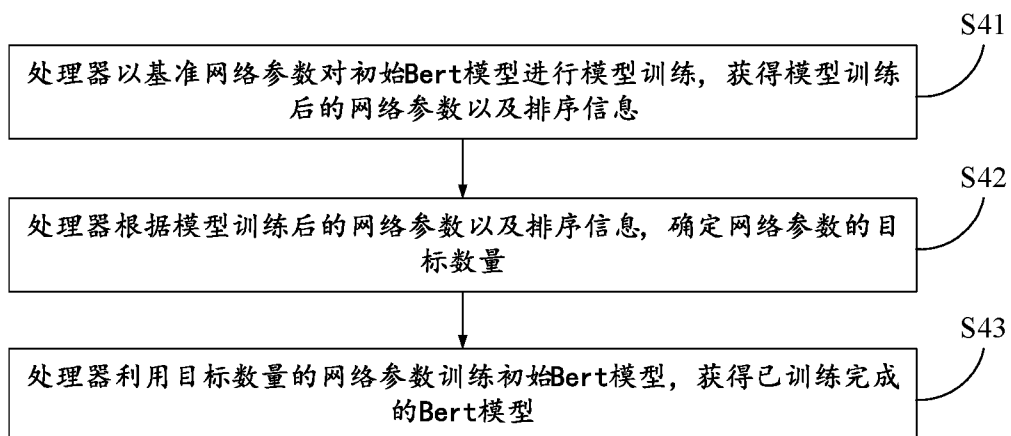


图 6

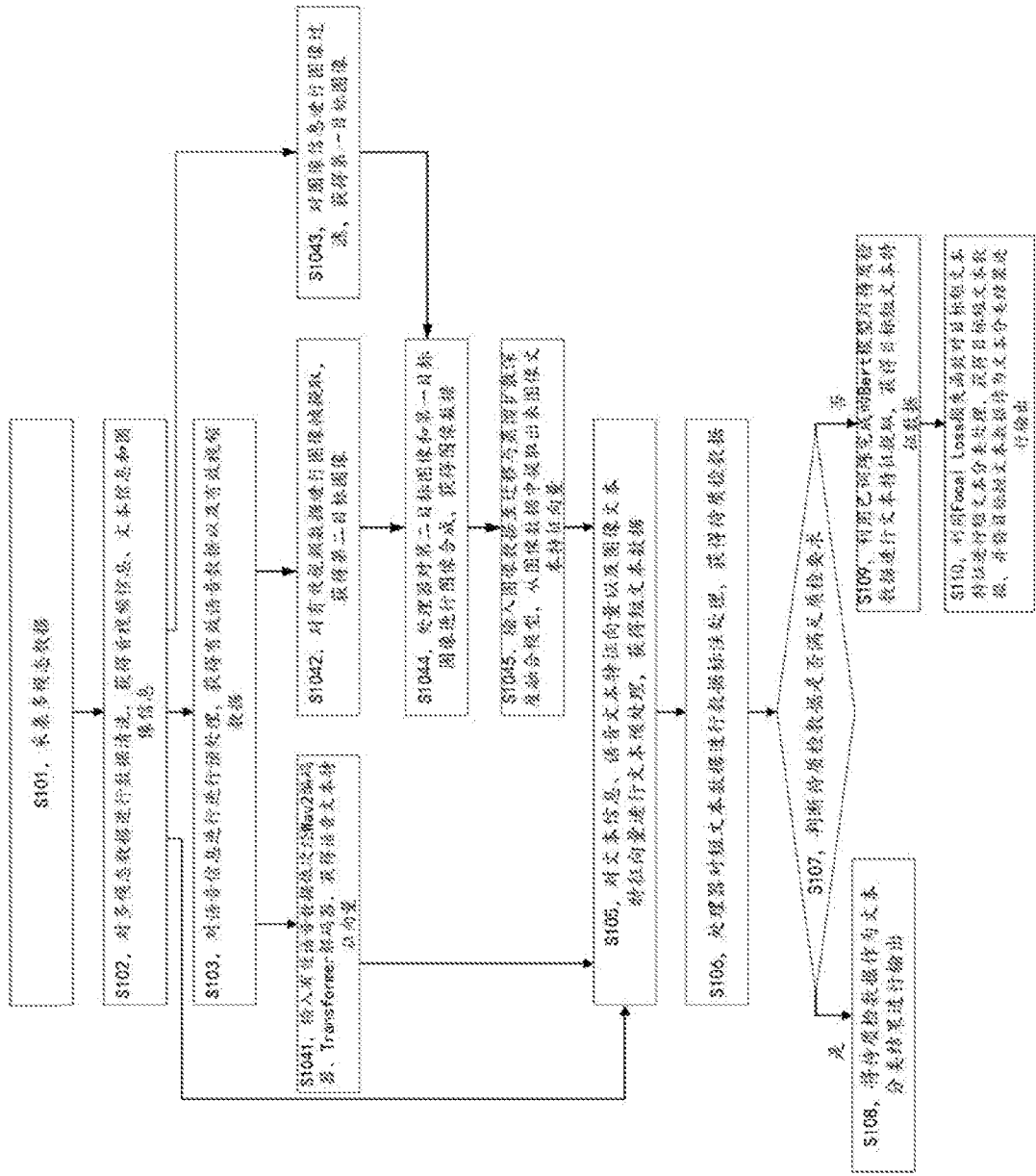


图 7

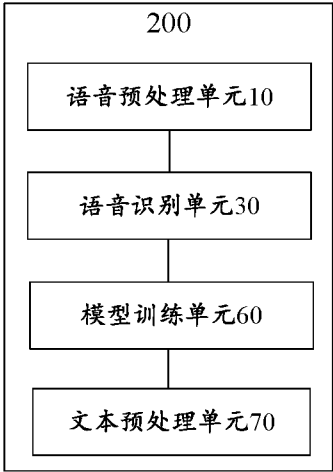


图 8

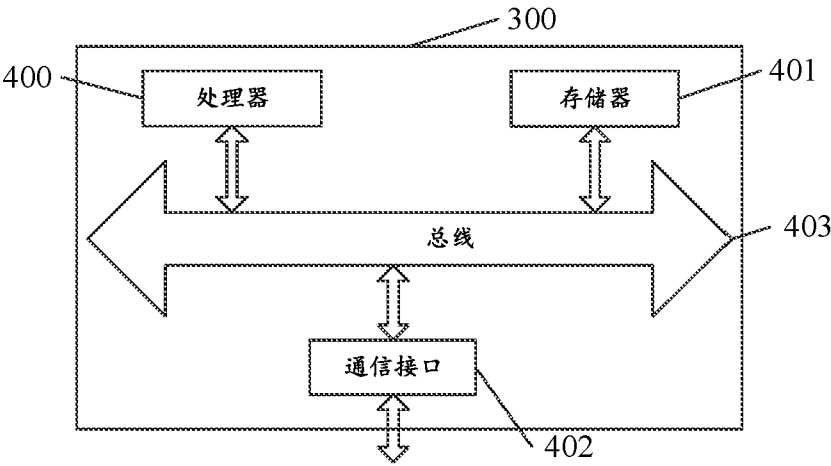


图 9

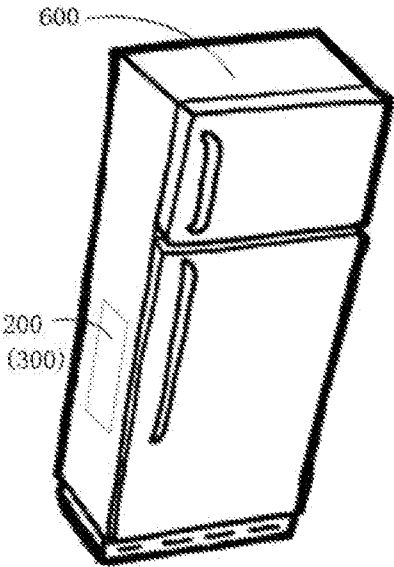


图 10

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2024/098340

A. CLASSIFICATION OF SUBJECT MATTER G06F 16/35(2019.01)i; G10L 13/08(2013.01)i; G06V 10/40(2022.01)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) IPC: G06F G10L G06V G06N 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) VEN, CNABS, CNTXT, WOTXT, EPTXT, USTXT, CNKI, IEEE: 多模态, 视频, 音频, 语音, 文本, 分类, 帧提取, 过滤, 融合, 合成, 排序, 目标数量, 参数, 编码器, 压缩, 裁剪, 剪枝, 知识蒸馏, 层数, 权值, 权重, multi, modal, mode, video, audio, speech, text, classification, frame, filter, fuse, rank, encoder, transformer, BERT, compression, pruning, knowledge distillation, layer, 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 116910244 A (QINGDAO HAIER REFRIGERATOR CO., LTD. et al.) 20 October 2023 (2023-10-20) claims 1-16, and description, paragraph 0136	1-18
PX	CN 116956209 A (QINGDAO HAIER REFRIGERATOR CO., LTD. et al.) 27 October 2023 (2023-10-27) description, paragraphs 0051-0134	1-6, 11-12, 14-18
X	CN 116186258 A (QINGDAO HAIER REFRIGERATOR CO., LTD. et al.) 30 May 2023 (2023-05-30) description, paragraphs 0035-0128	1-6, 11-12, 14-18
Y	CN 116186258 A (QINGDAO HAIER REFRIGERATOR CO., LTD. et al.) 30 May 2023 (2023-05-30) description, paragraphs 0035-0128	7-10, 13
☒ 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 05 September 2024		Date of mailing of the international search report 12 September 2024
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		Authorized officer Telephone No.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 115269774 A (QINGDAO HAIER TECHNOLOGY CO., LTD. et al.) 01 November 2022 (2022-11-01) description, paragraphs 0023-0111	7,13
Y	刘欢 等 (LIU, Huan et al.). "BERT模型的主要优化改进方法研究综 (A Review on Main Optimization Methods of BERT) " 数据分析与知识发现 (Data Analysis and Knowledge Discovery), Vol. 05, No. 01, 31 January 2021 (2021-01-31), pages 8-10	8-10
A	CN 116108176 A (QINGDAO HAIER REFRIGERATOR CO., LTD. et al.) 12 May 2023 (2023-05-12) entire document	1-18
A	CN 111079623 A (CHENGDU RUIYAN TECHNOLOGY CO., LTD.) 28 April 2020 (2020-04-28) entire document	1-18
A	US 2022327809 A1 (BEIJING BAIDU NETCOM SCIENCE TECHNOLOGY CO., LTD.) 13 October 2022 (2022-10-13) entire document	1-18

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2024/098340

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	116910244	A	20 October 2023	None			
CN	116956209	A	27 October 2023	None			
CN	116186258	A	30 May 2023	None			
CN	115269774	A	01 November 2022	WO	2024001101	A1	04 January 2024
CN	116108176	A	12 May 2023	WO	2024140430	A1	04 July 2024
CN	111079623	A	28 April 2020	None			
US	2022327809	A1	13 October 2022	JP	2022137145	A	21 September 2022
				JP	7429734	B2	08 February 2024
				CN	113240056	A	10 August 2021

A.主题的分类

G06F 16/35(2019.01)i; G10L 13/08(2013.01)i; G06V 10/40(2022.01)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)

IPC: G06F G10L G06V G06N

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))

VEN, CNABS, CNTXT, WOTXT, EPTXT, USTXT, CNKI, IEEE: 多模态, 视频, 音频, 语音, 文本, 分类, 帧提取, 过滤, 融合, 合成, 排序, 目标数量, 参数, 编码器, 压缩, 裁剪, 剪枝, 知识蒸馏, 层数, 权值, 权重, multi, modal, mode, video, audio, speech, text, classification, frame, filter, fuse, rank, encoder, transformer, BERT, compression, pruning, knowledge distillation, layer, 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 116910244 A (青岛海尔电冰箱有限公司 等) 2023年10月20日 (2023 - 10 - 20) 权利要求1-16, 说明书第0136段	1-18
PX	CN 116956209 A (青岛海尔电冰箱有限公司 等) 2023年10月27日 (2023 - 10 - 27) 说明书第0051-0134段	1-6,11-12,14-18
X	CN 116186258 A (青岛海尔电冰箱有限公司 等) 2023年5月30日 (2023 - 05 - 30) 说明书第0035-0128段	1-6,11-12,14-18
Y	CN 116186258 A (青岛海尔电冰箱有限公司 等) 2023年5月30日 (2023 - 05 - 30) 说明书第0035-0128段	7-10,13
Y	CN 115269774 A (青岛海尔科技有限公司 等) 2022年11月1日 (2022 - 11 - 01) 说明书第0023-0111段	7,13
Y	刘欢 等. "BERT模型的主要优化改进方法研究综述" 数据分析与知识发现, 第05卷, 第01期, 2021年1月31日 (2021 - 01 - 31), 第8-10页	8-10

☒ 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

2024年9月5日

国际检索报告邮寄日期

2024年9月12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

王晓燕

电话号码 (+86) 010-53961384

PCT/ISA/210 表(第2页) (2022年7月)

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 116108176 A (青岛海尔电冰箱有限公司 等) 2023年5月12日 (2023 - 05 - 12) 全文	1-18
A	CN 111079623 A (成都睿沿科技有限公司) 2020年4月28日 (2020 - 04 - 28) 全文	1-18
A	US 2022327809 A1 (BEIJING BAIDU NETCOM SCIENCE TECHNOLOGY CO., LTD.) 2022年10月13日 (2022 - 10 - 13) 全文	1-18

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2024/098340

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	116910244	A	2023年10月20日	无			
CN	116956209	A	2023年10月27日	无			
CN	116186258	A	2023年5月30日	无			
CN	115269774	A	2022年11月1日	WO	2024001101	A1	2024年1月4日
CN	116108176	A	2023年5月12日	WO	2024140430	A1	2024年7月4日
CN	111079623	A	2020年4月28日	无			
US	2022327809	A1	2022年10月13日	JP	2022137145	A	2022年9月21日
				JP	7429734	B2	2024年2月8日
				CN	113240056	A	2021年8月10日