



(12) 发明专利

(10) 授权公告号 CN 113743111 B

(45) 授权公告日 2024.06.04

(21) 申请号 202010865079.X

G06N 3/084 (2023.01)

(22) 申请日 2020.08.25

G06Q 40/03 (2023.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 113743111 A

(56) 对比文件

CN 108491380 A, 2018.09.04

CN 108920460 A, 2018.11.30

CN 110334814 A, 2019.10.15

CN 110728298 A, 2020.01.24

CN 111209738 A, 2020.05.29

CN 111353533 A, 2020.06.30

CN 111444721 A, 2020.07.24

US 2019147298 A1, 2019.05.16

US 2020167419 A1, 2020.05.28

US 2020251100 A1, 2020.08.06

WO 2019095572 A1, 2019.05.23

(43) 申请公布日 2021.12.03

(73) 专利权人 国家计算机网络与信息安全管理中心

地址 100029 北京市朝阳区裕民路甲3号

(72) 发明人 郭舒 陈桢豫 王丽宏 贺敏
毛乾任 李晨 钟盛海 黄洪仁

(74) 专利代理机构 北京华夏泰和知识产权代理有限公司 11662

专利代理师 韩来兵 李雪

(51) Int. Cl.

G06F 40/289 (2020.01)

G06F 40/30 (2020.01)

G06F 18/2415 (2023.01)

G06F 18/2431 (2023.01)

G06N 3/045 (2023.01)

王杰.基于多任务学习的情感分析技术研究.中国优秀硕士学位论文全文数据库 信息科技辑.2020,(第01期),1138-2757. (续)

审查员 廖雯雯

权利要求书4页 说明书18页 附图7页

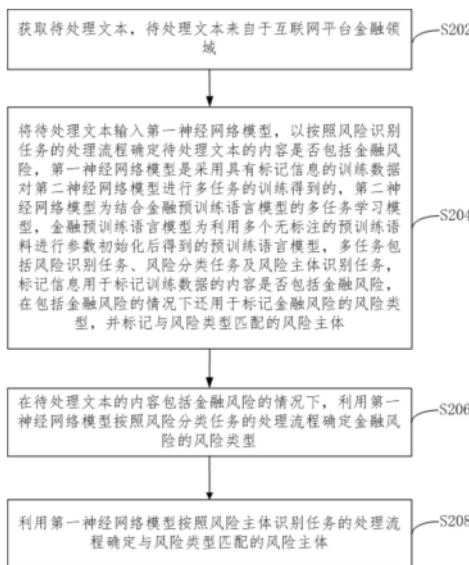
(54) 发明名称

基于文本预训练和多任务学习的金融风险预测方法及装置

(57) 摘要

本申请涉及一种基于文本预训练和多任务学习的金融风险预测方法及装置。该方法包括：获取待处理文本；将待处理文本输入第一神经网络模型，以按照风险识别任务的流程确定待处理文本的内容是否包括金融风险；在待处理文本的内容包括金融风险的情况下，利用第一神经网络模型按照风险分类任务的流程确定金融风险的风险类型；利用第一神经网络模型按照风险主体识别任务的流程确定与风险类型匹配的风险主体。本申请通过预训练语言模型技术解决了缺乏对语义的深度挖掘导致模型性能不佳的问题，并采用多任务处理，解决了数据量有限而且任务之间无法进行信息共享导致模型

性能不佳的技术问题。



CN 113743111 B

[接上页]

(56) 对比文件

赵亮.多模态数据融合算法研究.中国博士学位论文全文数据库 信息科技辑.2018,(第12期),I138-51.

Wenliang Dai 等.Kungfupanda at SemEval-2020 task 12:BERT-based multi-task learning for offensive language detection.arXiv.2020,1-7.

1. 一种基于文本预训练和多任务学习的金融风险预测方法,其特征在于,包括:
获取待处理文本,其中,所述待处理文本来自于互联网平台金融领域;

将所述待处理文本输入第一神经网络模型,以按照风险识别任务的处理流程确定所述待处理文本的内容是否包括金融风险,其中,所述第一神经网络模型是采用具有标记信息的训练数据对第二神经网络模型进行多任务的训练得到的,所述第二神经网络模型为结合金融预训练语言模型的多任务学习模型,所述金融预训练语言模型为利用多个无标注的预训练语料进行参数初始化后得到的预训练语言模型,所述多任务包括所述风险识别任务、风险分类任务及风险主体识别任务,所述标记信息用于标记所述训练数据的内容是否包括所述金融风险,在包括所述金融风险的情况下还用于标记所述金融风险的风险类型,并标记与所述风险类型匹配的风险主体;

在所述待处理文本的内容包括所述金融风险的情况下,利用所述第一神经网络模型按照所述风险分类任务的处理流程确定所述金融风险的风险类型;

利用所述第一神经网络模型按照所述风险主体识别任务的处理流程确定与所述风险类型匹配的风险主体。

2. 根据权利要求1所述的方法,其特征在于,将所述待处理文本输入第一神经网络模型之前,所述方法还包括按照如下方式对所述第二神经网络模型进行所述多任务的训练,得到所述第一神经网络模型:

从训练数据池中随机确定一个批次的所述训练数据,其中,所述训练数据包括用于所述风险识别任务、所述风险分类任务及所述风险主体识别任务的训练数据;

将所述训练数据输入所述第二神经网络模型,并在所述第二神经网络模型的预训练参数的基础上继续训练所述第二神经网络模型的各项参数;

采用早停的训练方式,在所述第二神经网络模型对测试数据的识别准确度达到最优值的情况下,将所述第二神经网络模型作为所述第一神经网络模型;

在所述第二神经网络模型对所述测试数据的识别准确度未达到所述最优值的情况下,继续使用所述训练数据对所述第二神经网络模型进行训练,以调整所述第二神经网络模型内各网络层中的参数的数值,直至所述第二神经网络模型对所述测试数据的识别准确度达到所述最优值。

3. 根据权利要求2所述的方法,其特征在于,将所述训练数据输入所述第二神经网络模型之前,所述方法还包括按照如下方式利用无标注的所述预训练语料对深度双向语言模型进行预训练,得到所述金融预训练语言模型:

获取所述预训练语料,其中,所述预训练语料来自于互联网平台金融领域;

按照第一预训练语言模型的输入要求对所述预训练语料进行预处理,其中,所述第一预训练语言模型为所述深度双向语言模型,所述第一预训练语言模型为基于通用领域语料进行预训练得到的预训练语言模型;

利用经过预处理的所述预训练语料对所述第一预训练语言模型进行预训练;

在所述第一预训练语言模型在目标预训练任务上的性能达到目标性能阈值的情况下,将所述第一预训练语言模型作为所述金融预训练语言模型;

在所述第一预训练语言模型在所述目标预训练任务上的性能未达到最优值的情况下,继续使用所述预训练语料对所述第一预训练语言模型进行预训练,以调整所述第一预训练

语言模型内各网络层中的参数的数值,直至所述第一预训练语言模型在所述目标预训练任务上的性能达到最优值。

4. 根据权利要求3所述的方法,其特征在于,将所述训练数据输入所述第二神经网络模型之前,所述方法还包括按照如下方式结合所述金融预训练语言模型,得到所述第二神经网络模型:

在所述金融预训练语言模型的输出层分别添加用于所述风险识别任务、所述风险分类任务及所述风险主体识别任务的输出层,得到所述第二神经网络模型。

5. 根据权利要求2所述的方法,其特征在于,从训练数据池中随机确定一个批次的所述训练数据之前,所述方法还包括按照如下方式构建所述训练数据池:

按照预设的每个批次的的数据量大小将用于所述风险识别任务、所述风险分类任务及所述风险主体识别任务的所述训练数据划分为多个批次;

将所有批次的所述训练数据进行无序混合,得到所述训练数据池。

6. 根据权利要求2所述的方法,其特征在于,在所述第二神经网络模型的预训练参数的基础上继续训练所述第二神经网络模型的各项参数包括:

将所述第二神经网络模型的嵌入层和编码层作为共享参数区,将所述第二神经网络模型的各个输出层分别作为私有参数区,其中,所述私有参数区包括第一私有参数区、第二私有参数区及第三私有参数区,所述第一私有参数区为所述风险识别任务的输出层,所述第二私有参数区为所述风险分类任务的输出层,所述第三私有参数区为所述风险主体识别任务的输出层;

将所述第一私有参数区、所述第二私有参数区及所述第三私有参数区的学习率固定为第一学习率,并利用所述训练数据对所述第二神经网络模型进行训练,以在多个第二学习率中确定所述共享参数区的第一目标学习率,其中,所述第一目标学习率为适用于所述共享参数区的最优学习率;

将所述共享参数区的学习率固定为所述第一目标学习率,并利用所述训练数据对所述第二神经网络模型进行训练,以在目标范围内分别确定所述第一私有参数区、所述第二私有参数区及所述第三私有参数区的第二目标学习率,其中,所述第二目标学习率为分别适用于所述第一私有参数区、所述第二私有参数区及所述第三私有参数区的最优学习率。

7. 根据权利要求6所述的方法,其特征在于,在所述第二神经网络模型的预训练参数的基础上继续训练所述第二神经网络模型的各项参数还包括:

通过所述共享参数区的参数共享,在训练所述第二神经网络模型的过程中确定目标隐层参数,其中,所述目标隐层参数为同时适用于所述第一私有参数区、所述第二私有参数区及所述第三私有参数区的隐层参数。

8. 根据权利要求2所述的方法,其特征在于,在所述第二神经网络模型对测试数据的识别准确度达到最优值的情况下,将所述第二神经网络模型作为所述第一神经网络模型包括:

获取第一测试数据;

将所述第一测试数据输入所述第二神经网络模型,以按照所述风险识别任务的处理流程对所述第一测试数据进行处理,得到所述风险识别任务的输出层输出的风险识别结果;

确定所述风险识别结果的精确率与召回率的第一调和平均值,并筛选出第二测试数

据,其中,所述第二测试数据为所述标记信息为有风险且所述风险识别结果为有风险的所述第一测试数据;

利用所述第二神经网络模型按照所述风险分类任务的处理流程对所述第二测试数据进行处理,得到所述风险分类任务的输出层输出的风险分类结果;

确定所述风险分类结果的准确率和所述风险分类结果的排序倒数值,并筛选出第三测试数据,其中,所述第三测试数据为所述风险分类结果与所述标记信息标记的风险类型匹配的所述第二测试数据;

利用所述第二神经网络模型按照所述风险主体识别任务的处理流程对所述第三测试数据进行处理,得到所述风险主体识别任务的输出层输出的风险主体识别结果;

确定所述风险主体识别结果的完全匹配值和所述风险主体识别结果的精确率与召回率的第二调和平均值;

在所述第一调和平均值、所述准确率、所述排序倒数值、所述完全匹配值以及所述第二调和平均值达到对应的预设指标的情况下,将所述第二神经网络模型确定为所述第一神经网络模型。

9. 根据权利要求1至8任一所述的方法,其特征在于,所述风险识别任务的处理流程包括:

按照预设对应关系将所述待处理文本转换为第一标记序列;

将所述第一标记序列通过所述第一神经网络模型的嵌入层和编码层,得到所述待处理文本的第一语义表示向量,其中,所述第一语义表示向量为包含所述待处理文本上下文语义信息的向量;

对所述第一语义表示向量进行线性变换,得到第二语义表示向量,其中,所述第二语义表示向量是通过所述风险识别任务的输出层的私有参数进行处理后得到的;

采用Softmax分类的方式对所述第二语义表示向量进行处理,得到第一概率分布,其中,所述第一概率分布为所述待处理文本的内容包含所述金融风险的概率值,所述第一概率分布是通过所述风险识别任务的输出层得到的;

根据所述第一概率分布确定所述待处理文本是否包含所述金融风险。

10. 根据权利要求9所述的方法,其特征在于,所述风险分类任务的处理流程包括:

对所述第一语义表示向量进行线性变换,得到第三语义表示向量,其中,所述第三语义表示向量是通过所述风险分类任务的输出层的私有参数进行处理后得到的;

采用Softmax分类的方式对所述第三语义表示向量进行处理,得到第二概率分布,其中,所述第二概率分布为所述金融风险的所述风险类型属于各个类型的概率值,所述第二概率分布是通过所述风险分类任务的输出层得到的;

根据所述第二概率分布确定所述金融风险的所述风险类型。

11. 根据权利要求10所述的方法,其特征在于,所述风险主体识别任务的处理流程包括:

将所述待处理文本与所述金融风险的所述风险类型进行拼接,并按照所述预设对应关系转换为第二标记序列;

将所述第二标记序列通过所述第一神经网络模型的嵌入层和编码层,得到第四语义表示向量,其中,所述第四语义表示向量为包含拼接后的所述待处理文本和所述风险类型上

下文语义信息的向量；

对所述第四语义表示向量进行线性变换,得到第五语义表示向量,其中,所述第五语义表示向量是通过所述风险主体识别任务的输出层的私有参数进行处理后得到的；

利用所述第五语义表示向量确定第三概率分布和第四概率分布,其中,所述第三概率分布为所述第五语义表示向量中的每一个字向量作为与所述风险类型匹配的所述风险主体的起始字向量的概率值,所述第四概率分布为所述第五语义表示向量中的每一个字向量作为与所述风险类型匹配的所述风险主体的终止字向量的概率值；

根据所述第三概率分布和所述第四概率分布确定与所述风险类型匹配的所述风险主体。

12. 一种基于文本预训练和多任务学习的金融风险预测装置,其特征在于,包括:

获取模块,用于获取待处理文本,其中,所述待处理文本来自于互联网平台金融领域；

风险识别模块,用于将所述待处理文本输入第一神经网络模型,以按照风险识别任务的处理流程确定所述待处理文本的内容是否包括金融风险,其中,所述第一神经网络模型是采用具有标记信息的训练数据对第二神经网络模型进行多任务的训练得到的,所述第二神经网络模型为结合金融预训练语言模型的多任务学习模型,所述金融预训练语言模型为利用多个无标注的预训练语料进行参数初始化后得到的预训练语言模型,所述多任务包括所述风险识别任务、风险分类任务及风险主体识别任务,所述标记信息用于标记所述训练数据的内容是否包括所述金融风险,在包括所述金融风险的情况下还用于标记所述金融风险的风险类型,并标记与所述风险类型匹配的风险主体；

风险分类模块,用于在所述待处理文本的内容包括所述金融风险的情况下,利用所述第一神经网络模型按照所述风险分类任务的处理流程确定所述金融风险的风险类型；

风险主体识别模块,用于利用所述第一神经网络模型按照所述风险主体识别任务的处理流程确定与所述风险类型匹配的风险主体。

13. 一种计算机设备,包括存储器、处理器,所述存储器中存储有可在所述处理器上运行的计算机程序,其特征在于,所述处理器执行所述计算机程序时实现上述权利要求1至11任一项所述的方法的步骤。

14. 一种具有处理器可执行的非易失的程序代码的计算机可读介质,其特征在于,所述程序代码使所述处理器执行所述权利要求1至11任一所述方法。

基于文本预训练和多任务学习的金融风险预测方法及装置

技术领域

[0001] 本申请涉及风险预测技术领域,尤其涉及一种基于文本预训练和多任务学习的金融风险预测方法及装置。

背景技术

[0002] 随着互联网在金融业务上的深度应用,互联网金融逐渐走入大众视野,互联网金融指的是通过或依托互联网技术和工具进行资金融通和支付及相关信息服务等业务的行为,它利用互联网平台为金融提供了新的获取信息的方式,多样化的风险管理工具与风险分散工具。

[0003] 当前互联网+金融格局,由传统金融机构和非金融机构组成。传统金融机构主要为传统金融业务的互联网创新以及电商化创新、APP软件等;非金融机构则主要是指利用互联网技术进行金融运作的电商企业、(P2P)模式的网络借贷平台,众筹模式的网络投资平台,挖财类(模式)的手机理财APP(理财宝类),以及第三方支付平台等。

[0004] 从机构和市场指标的规模上来说,目前我国互联网金融似乎已经走到了全球前列,但仍然存在“短板”。互联网金融带给人们便利的同时,也在不断暴露出层出不穷的风险,如“P2P跑路”、“网络高利贷和暴力催收”等。互联网本身具有不分地域、信息涉众面广的特点,而这些使得互联网金融风险传播速度快,风险跨域处置难度大。另外,打击非法集资是防范金融风险的重要领域,当前非法集资形式依然严峻,新案高发与陈案积压并存,区域及行业风险集中,上网跨域特点明显,集资参与人量大面广。由此可见,对互联网金融风险的预警和防控已经势在必行,迫在眉睫。

[0005] 目前,相关技术中,对金融风险的预测都存在片面性。金融风险预测任务包含的研究问题种类十分广泛,一些研究专注于利用用户、公司或机构的特征判断其是否存在潜在的金融风险,即将金融风险预测视为一个二分类问题;另一些研究旨在判定具体目标的金融风险等级,即将其视为一个多分类问题;也有研究指向于预测公司或其他金融机构的金融风险评分,即将其视为一个回归问题。

[0006] 传统的金融风险预测任务一般采用量化的数据作为模型的输入,即传统金融风险预测的研究大多是基于量化数据进行,将样本的量化指标直接作为特征进行分类。比如,在预测欺诈行为的任务中可能采取用户的收入、存款数额等数据作为输入;而在预测银行破产中常采用银行目前的总资产、现金流、总贷款金额等。然而,量化数据具有数据量有限、非业内人士难以获取的特点,而对于互联网上大量存在且容易获得的金融文本数据,当前的金融风险研究对其利用依旧不够充分。

[0007] 另外,虽然针对金融文本数据进行分析从而进行金融风险预测的工作较少见,但针对金融文本数据进行其他方面研究分析的工作也常有出现。此类研究一般被统称为金融文本挖掘。金融文本挖掘的初衷是,利用文本挖掘技术对文本数据进行分析以进行更好的决策。目前,在金融领域的文本挖掘工作主要用于外汇汇率预测、股票市场预测、客户流失预测等,还有一些网络安全方面的应用,包括网络钓鱼检测、垃圾邮件检测和欺诈检测等。

金融领域的文本挖掘工作一般采用新闻标题或者新闻内容等文本数据,结合一些常见的机器学习算法(LR、SVM、DT、k-NN和NB等),对数据进行分类任务。现阶段常见的金融文本挖掘工作,譬如基于金融新闻的外汇汇率预测和股票市场预测等,大多在文本预处理和特征构建的方法采取比较简单的方法,一般采用词袋模型进行文本预处理,忽略位置关系和词汇之间的关联,得到的词向量比较稀疏;在特征构建阶段,普遍将词频作为特征,缺乏对语义的深度挖掘。

[0008] 此外,金融风险预测多采用单模型或者集成模型的方法,没有很好地解决某些场景下训练数据不充分的问题,不能充分利用任务之间的共享信息来优化提升各个任务的效果。

[0009] 针对上述的问题,目前尚未提出有效的解决方案。

发明内容

[0010] 本申请提供了一种基于文本预训练和多任务学习的金融风险预测方法及装置,以解决上述“特征构建比较简单,得到的词向量比较稀疏,缺乏对语义的深度挖掘导致模型性能不佳,数据量有限而且任务之间无法进行信息共享导致模型性能不佳”的技术问题。

[0011] 第一方面,本申请提供了一种基于文本预训练和多任务学习的金融风险预测方法,包括:获取待处理文本,待处理文本来自于互联网平台金融领域;将待处理文本输入第一神经网络模型,以按照风险识别任务的处理流程确定待处理文本的内容是否包括金融风险,第一神经网络模型是采用具有标记信息的训练数据对第二神经网络模型进行多任务的训练得到的,第二神经网络模型为结合金融预训练语言模型的多任务学习模型,金融预训练语言模型为利用多个无标注的预训练语料进行参数初始化后得到的预训练语言模型,多任务包括风险识别任务、风险分类任务及风险主体识别任务,标记信息用于标记训练数据的内容是否包括金融风险,在包括金融风险的情况下还用于标记金融风险的风险类型,并标记与风险类型匹配的风险主体;在待处理文本的内容包括金融风险的情况下,利用第一神经网络模型按照风险分类任务的处理流程确定金融风险的风险类型;利用第一神经网络模型按照风险主体识别任务的处理流程确定与风险类型匹配的风险主体。

[0012] 可选地,将待处理文本输入第一神经网络模型之前,该方法还包括按照如下方式对第二神经网络模型进行多任务的训练,得到第一神经网络模型:从训练数据池中随机确定一个批次的训练数据,训练数据包括用于风险识别任务、风险分类任务及风险主体识别任务的训练数据;将训练数据输入第二神经网络模型,并在第二神经网络模型的预训练参数的基础上继续训练第二神经网络模型的各项参数;采用早停的训练方式,在第二神经网络模型对测试数据的识别准确度达到最优值的情况下,将第二神经网络模型作为第一神经网络模型;在第二神经网络模型对测试数据的识别准确度未达到最优值的情况下,继续使用训练数据对第二神经网络模型进行训练,以调整第二神经网络模型内各网络层中的参数的数值,直至第二神经网络模型对测试数据的识别准确度达到最优值。

[0013] 可选地,将训练数据输入第二神经网络模型之前,该方法还包括按照如下方式利用无标注的预训练语料对深度神经网络模型进行预训练,得到金融预训练语言模型:获取预训练语料,预训练语料来自于互联网平台金融领域;按照第一预训练语言模型的输入要求对预训练语料进行预处理,第一预训练语言模型为深度神经网络模型,第一预训练语言

模型为基于通用领域语料进行预训练得到的预训练语言模型；利用经过预处理的预训练语料对第一预训练语言模型进行预训练；在第一预训练语言模型在目标预训练任务上的性能达到目标性能阈值的情况下，将第一预训练语言模型作为金融预训练语言模型；在第一预训练语言模型在目标预训练任务上的性能未达到目标性能阈值的情况下，继续使用预训练语料对第一预训练语言模型进行预训练，以调整第一预训练语言模型内各网络层中的参数的数值，直至第一预训练语言模型在目标预训练任务上的性能达到目标性能阈值。

[0014] 可选地，将训练数据输入第二神经网络模型之前，该方法还包括按照如下方式结合金融预训练语言模型，得到第二神经网络模型：在金融预训练语言模型的输出层分别添加用于风险识别任务、风险分类任务及风险主体识别任务的输出层，得到第二神经网络模型。

[0015] 可选地，从训练数据池中随机确定一个批次的训练数据之前，该方法还包括按照如下方式构建训练数据池：按照预设的每个批次的的数据量大小将用于风险识别任务、风险分类任务及风险主体识别任务的训练数据划分为多个批次；将所有批次的训练数据进行无序混合，得到训练数据池。

[0016] 可选地，在第二神经网络模型的预训练参数的基础上继续训练第二神经网络模型的各项参数包括：将第二神经网络模型的嵌入层和编码层作为共享参数区，将第二神经网络模型的各个输出层分别作为私有参数区，私有参数区包括第一私有参数区、第二私有参数区及第三私有参数区，第一私有参数区为风险识别任务的输出层，第二私有参数区为风险分类任务的输出层，第三私有参数区为风险主体识别任务的输出层；将第一私有参数区、第二私有参数区及第三私有参数区的学习率固定为第一学习率，并利用训练数据对第二神经网络模型进行训练，以在多个第二学习率中确定共享参数区的第一目标学习率，第一目标学习率为适用于共享参数区的最优学习率；将共享参数区的学习率固定为第一目标学习率，并利用训练数据对第二神经网络模型进行训练，以在目标范围内分别确定第一私有参数区、第二私有参数区及第三私有参数区的第二目标学习率，第二目标学习率为分别适用于第一私有参数区、第二私有参数区及第三私有参数区的最优学习率。

[0017] 可选地，在第二神经网络模型的预训练参数的基础上继续训练第二神经网络模型的各项参数还包括：通过共享参数区的参数共享，在训练第二神经网络模型的过程中确定目标隐层参数，目标隐层参数为同时适用于第一私有参数区、第二私有参数区及第三私有参数区的隐层参数。

[0018] 可选地，在第二神经网络模型对测试数据的识别准确度达到最优值的情况下，将第二神经网络模型作为第一神经网络模型包括：获取第一测试数据；将第一测试数据输入第二神经网络模型，以按照风险识别任务的处理流程对第一测试数据进行处理，得到风险识别任务的输出层输出的风险识别结果；确定风险识别结果的精确率与召回率的第一调和平均值，并筛选出第二测试数据，第二测试数据为标记信息为有风险且风险识别结果为有风险的第一测试数据；利用第二神经网络模型按照风险分类任务的处理流程对第二测试数据进行处理，得到风险分类任务的输出层输出的风险分类结果；确定风险分类结果的准确率和风险分类结果的排序倒数值，并筛选出第三测试数据，第三测试数据为风险分类结果与标记信息标记的风险类型匹配的第二测试数据；利用第二神经网络模型按照风险主体识别任务的处理流程对第三测试数据进行处理，得到风险主体识别任务的输出层输出的风险

主体识别结果;确定风险主体识别结果的完全匹配值和风险主体识别结果的精确率与召回率的第二调和平均值;在第一调和平均值、准确率、排序倒数、完全匹配值以及第二调和平均值达到对应的预设指标的情况下,将第二神经网络模型确定为所述第一神经网络模型。

[0019] 可选地,风险识别任务的处理流程包括:按照预设对应关系将待处理文本转换为第一标记序列;将第一标记序列通过第一神经网络模型的嵌入层和编码层,得到所述待处理文本的第一语义表示向量,第一语义表示向量为包含待处理文本上下文语义信息的向量;对第一语义表示向量进行线性变换,得到第二语义表示向量,第二语义表示向量是通过风险识别任务的输出层的私有参数进行处理后得到的;采用Softmax分类的方式对第二语义表示向量进行处理,得到第一概率分布,第一概率分布为待处理文本的内容包含金融风险的概率值,第一概率分布是通过风险识别任务的输出层得到的;根据第一概率分布确定待处理文本是否包含金融风险。

[0020] 可选地,风险分类任务的处理流程包括:对第一语义表示向量进行线性变换,得到第三语义表示向量,第三语义表示向量是通过风险分类任务的输出层的私有参数进行处理后得到的;采用Softmax分类的方式对第三语义表示向量进行处理,得到第二概率分布,第二概率分布为金融风险的风险类型属于各个类型的概率值,第二概率分布是通过风险分类任务的输出层得到的;根据第二概率分布确定金融风险的风险类型。

[0021] 可选地,风险主体识别任务的处理流程包括:将待处理文本与金融风险的风险类型进行拼接,并按照预设对应关系转换为第二标记序列;将第二标记序列通过第一神经网络模型的嵌入层和编码层,得到第四语义表示向量,第四语义表示向量为包含拼接后的待处理文本和风险类型上下文语义信息的向量;对第四语义表示向量进行线性变换,得到第五语义表示向量,第五语义表示向量是通过风险主体识别任务的输出层的私有参数进行处理后得到的;利用第五语义表示向量确定第三概率分布和第四概率分布,第三概率分布为第五语义表示向量中的每一个字向量作为与风险类型匹配的风险主体的起始字向量的概率值,第四概率分布为第五语义表示向量中的每一个字向量作为与风险类型匹配的风险主体的终止字向量的概率值;根据第三概率分布和所述第四概率分布确定与风险类型匹配的风险主体。

[0022] 第二方面,本申请提供了一种基于文本预训练和多任务学习的金融风险预测装置,包括:获取模块,用于获取待处理文本,待处理文本来自于互联网平台金融领域;风险识别模块,用于将待处理文本输入第一神经网络模型,以按照风险识别任务的处理流程确定待处理文本的内容是否包括金融风险,第一神经网络模型是采用具有标记信息的训练数据对第二神经网络模型进行多任务的训练得到的,第二神经网络模型为结合金融预训练语言模型的多任务学习模型,金融预训练语言模型为利用多个无标注的预训练语料进行参数初始化后得到的预训练语言模型,多任务包括风险识别任务、风险分类任务及风险主体识别任务,标记信息用于标记训练数据的内容是否包括金融风险,在包括金融风险的情况下还用于标记金融风险的风险类型,并标记与风险类型匹配的风险主体;风险分类模块,用于在待处理文本的内容包括金融风险的情况下,利用第一神经网络模型按照风险分类任务的处理流程确定金融风险的风险类型;风险主体识别模块,用于利用第一神经网络模型按照风险主体识别任务的处理流程确定与风险类型匹配的风险主体。

[0023] 第三方面,本申请提供了一种计算机设备,包括存储器、处理器,存储器中存储有可在处理器上运行的计算机程序,处理器执行计算机程序时实现上述第一方面任一项方法的步骤。

[0024] 第四方面,本申请还提供了一种具有处理器可执行的非易失的程序代码的计算机可读介质,程序代码使处理器执行上述第一方面任一方法。

[0025] 本申请实施例提供的上述技术方案与相关技术相比具有如下优点:

[0026] 本申请利用预训练语言模型技术,解决了缺乏对语义的深度挖掘导致模型性能不佳的问题,并采用多任务处理,解决了数据量有限而且任务之间无法进行信息共享导致模型性能不佳的技术问题。此外,多任务处理的方法降低了模型参数总量,具有节约存储空间并提升模型加载运行速度的效果。

附图说明

[0027] 此处的附图被并入说明书中并构成本说明书的一部分,示出了符合本申请的实施例,并与说明书一起用于解释本申请的原理。

[0028] 为了更清楚地说明本申请实施例或相关技术中的技术方案,下面将对实施例或相关技术描述中所需要使用的附图作简单地介绍,显而易见地,对于本领域普通技术人员而言,在不付出创造性劳动性的前提下,还可以根据这些附图获得其他的附图。

[0029] 图1为根据本申请实施例提供的一种可选的基于文本预训练和多任务学习的金融风险预测方法硬件环境示意图;

[0030] 图2为根据本申请实施例提供的一种可选的基于文本预训练和多任务学习的金融风险预测方法流程图;

[0031] 图3为根据本申请实施例提供的一种可选的多任务学习的模型训练流程图;

[0032] 图4为根据本申请实施例提供的一种可选的金融预训练语言模型训练流程图;

[0033] 图5为根据本申请实施例提供的一种可选的训练数据池构建流程图;

[0034] 图6为根据本申请实施例提供的一种可选的参数训练流程图;

[0035] 图7为根据本申请实施例提供的一种可选的串行模型测试流程图;

[0036] 图8为根据本申请实施例提供的一种可选的基于文本预训练和多任务学习的金融风险预测装置框图。

具体实施方式

[0037] 为使本申请实施例的目的、技术方案和优点更加清楚,下面将结合本申请实施例中的附图,对本申请实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例是本申请的一部分实施例,而不是全部的实施例。基于本申请中的实施例,本领域普通技术人员在没有做出创造性劳动的前提下所获得的所有其他实施例,都属于本申请保护的范围。

[0038] 在后续的描述中,使用用于表示元件的诸如“模块”、“部件”或“单元”的后缀仅为为了有利于本申请的说明,其本身并没有特定的意义。因此,“模块”与“部件”可以混合地使用。

[0039] 相关技术中,对金融风险的预测都存在片面性。金融风险预测任务包含的研究问题种类十分广泛,一些研究专注于利用用户、公司或机构的特征判断其是否存在潜在的金

融风险,即将金融风险预测视为一个二分类问题;另一些研究旨在判定具体目标的金融风险等级,即将其视为一个多分类问题;也有研究指向于预测公司或其他金融机构的金融风险评分,即将其视为一个回归问题。

[0040] 传统金融风险预测的研究大多是基于量化数据进行,将样本的量化指标直接作为特征进行分类。然而,量化数据具有数据量有限、非业内人士难以获取的特点。对目前的金融风险研究对互联网上大量存在且容易获得的金融文本数据的利用不够充分。

[0041] 现阶段常见的金融文本挖掘工作,譬如基于金融新闻的外汇汇率预测和股票市场预测等,大多在文本预处理和特征构建的方法采取比较简单的方法,一般采用词袋模型进行文本预处理,忽略位置关系和词汇之间的关联,得到的词向量比较稀疏;在特征构建阶段,普遍将词频作为特征,缺乏对语义的深度挖掘。能够利用目前比较先进的自然语言处理技术,深度挖掘文本中蕴含的语义特征信息的工作非常少见。

[0042] 此外,金融风险预测多采用单模型或者集成模型的方法,没有很好地解决某些场景下训练数据不充分的问题,不能充分利用任务之间的共享信息来优化提升各个任务的效果。

[0043] 为了解决背景技术中提及的问题,根据本申请实施例的一方面,提供了一种基于文本预训练和多任务学习的金融风险预测方法的实施例。

[0044] 可选地,在本申请实施例中,上述方法可以应用于如图1所示的由终端101和服务器103所构成的硬件环境中。如图1所示,服务器103通过网络与终端101进行连接,可用于为终端或终端上安装的客户端提供服务,可在服务器上或独立于服务器设置数据库105,用于为服务器103提供数据存储服务,上述网络包括但不限于:广域网、城域网或局域网,终端101包括但不限于PC、手机、平板电脑等。

[0045] 本申请实施例中的一种基于文本预训练和多任务学习的金融风险预测方法可以由服务器103来执行,还可以是由服务器103和终端101共同执行,如图2所示,该方法可以包括以下步骤:

[0046] 步骤S202,获取待处理文本,待处理文本来自于互联网平台金融领域。

[0047] 本申请实施例中,上述的待处理文本可以是英文、中文等语言的文本,上述互联网平台可以是财经新闻网站等金融领域的互联网平台。

[0048] 步骤S204,将待处理文本输入第一神经网络模型,以按照风险识别任务的处理流程确定待处理文本的内容是否包括金融风险,第一神经网络模型是采用具有标记信息的训练数据对第二神经网络模型进行多任务的训练得到的,第二神经网络模型为结合金融预训练语言模型的多任务学习模型,金融预训练语言模型为利用多个无标注的预训练语料进行参数初始化后得到的预训练语言模型,多任务包括风险识别任务、风险分类任务及风险主体识别任务,标记信息用于标记训练数据的内容是否包括金融风险,在包括金融风险的情况下还用于标记金融风险的风险类型,并标记与风险类型匹配的风险主体。

[0049] 本申请实施例中,基于预训练语言模型训练本申请使用的神经网络模型,可以提取到输入文本表示的上下文关联信息,即能够挖掘文本的深度语义特征。预训练语言模型是利用无标注的预训练语料进行训练后得到的,其中,本申请中的第二神经网络模型是建立在金融预训练语言模型基础之上,在金融预训练语言模型的输出层增加用于多任务输出的输出层得到的多任务学习模型,金融预训练语言模型是建立在通用领域预训练语言模型

的基础上采用未标注的金融文本语料进行训练得到的预训练语言模型。

[0050] 本申请实施例中,上述标记信息至少标识出训练数据中是否含有金融风险,还可以在具有金融风险的训练数据中标记出该金融风险的类型,如“重组失败”、“投诉维权”、“评级调整”等,并标记出与该风险类型匹配的主体。

[0051] 本申请实施例中,可以识别待处理文本是否包含金融风险。例如,待处理文本为:“辉丰股份副总因涉嫌污染环境案被监视居住”,显然,该文本中包含对“辉丰股份”公司的负面信息,有一定的潜在金融风险存在,模型对其识别的结果应为“有风险”。金融风险识别任务是一个文本二分类任务。

[0052] 步骤S206,在待处理文本的内容包括金融风险的情况下,利用第一神经网络模型按照风险分类任务的处理流程确定金融风险的风险类型。

[0053] 本申请实施例中,可以对前一步已经判断为“有风险”的数据进行风险类型的定性分类。本申请实施例中,对于不同的数据集,可以建立不同的分类体系。优选地,本申请实施例建立的风险类型具体包括“重组失败”、“投诉维权”、“评级调整”、“涉嫌跑路”、“涉嫌集资”、“涉嫌诈骗”、“实控人股东变更”等。例如“辉丰股份副总因涉嫌污染环境案被监视居住”中包含的金融风险类型应为“涉嫌违法”。金融风险分类是一个文本多分类问题。

[0054] 步骤S208,利用第一神经网络模型按照风险主体识别任务的处理流程确定与风险类型匹配的风险主体。

[0055] 本申请实施例中,可以对前一步中已经判定金融风险类型的数据,进一步在文本中根据其风险类型抽取出对应的风险主体。例如“辉丰股份副总因涉嫌污染环境案被监视居住”中包含的金融风险类型为“涉嫌违法”,与“涉嫌违法”所匹配的金融风险主体应为“辉丰股份”。

[0056] 本申请实施例中,在模型训练、测试及实际应用时都采用多任务学习的方法,即风险识别任务、风险分类任务及风险主体识别任务这三个任务之间进行信息互通和共享,相比于对三个任务分别采用独立模型结构的设计,多任务学习结构能够将整个系统的模型参数降低到接近到原来的三分之一,更加轻量化和精简。并且经过实验表明,相较于为每个任务单独设置模型,多任务学习模型架构性能对每一个任务的性能都有显著提升。

[0057] 可选地,上述风险识别任务的处理流程可以包括以下步骤:

[0058] 步骤1,按照预设对应关系将待处理文本转换为第一标记序列。

[0059] 上述预设对应关系可以是来自记录有汉字(或单词)与标记编号之间的对应关系的词表中。

[0060] 步骤2,将第一标记序列通过第一神经网络模型的嵌入层和编码层,得到所述待处理文本的第一语义表示向量,第一语义表示向量为包含待处理文本上下文语义信息的向量。将标记序列转换为向量可以是采用自然语言处理技术(NLP)中的Word Embedding方法,例如one-hot编码等。

[0061] 步骤3,对第一语义表示向量进行线性变换,得到第二语义表示向量,第二语义表示向量是通过风险识别任务的输出层的私有参数进行处理后得到的。

[0062] 在风险识别任务中进行的线性变换是通过风险识别任务的输出层的私有参数进行处理后得到的,使得线性变换的结果能够更加适合于风险识别。

[0063] 步骤4,采用Softmax分类的方式对第二语义表示向量进行处理,得到第一概率分

布,第一概率分布为待处理文本的内容包含金融风险的概率值,第一概率分布是通过风险识别任务的输出层得到的。

[0064] Softmax用于最小化估计分类概率和“真实”分布之间的交叉熵,得到的是归一化概率,可用于本申请实施例中的风险识别任务的二分类问题。

[0065] 步骤5,根据第一概率分布确定待处理文本是否包含金融风险。

[0066] 可以根据实际需要设定划分阈值,例如第一概率分布中待处理文本的内容有金融风险的概率为0.7,则划定该待处理文本的内容包含金融风险。

[0067] 可选地,上述风险分类任务的处理流程可以包括以下步骤:

[0068] 步骤1,对第一语义表示向量进行线性变换,得到第三语义表示向量,第三语义表示向量是通过风险分类任务的输出层的私有参数进行处理后得到的。

[0069] 在风险分类任务中进行的线性变换是通过风险分类任务的输出层的私有参数进行处理后得到的,使得线性变换的结果能够更加适合于风险分类。

[0070] 步骤2,采用Softmax分类的方式对第三语义表示向量进行处理,得到第二概率分布,第二概率分布为金融风险的风险类型属于各个类型的概率值,第二概率分布是通过风险分类任务的输出层得到的。

[0071] Softmax也可以用于本申请实施例中的风险分类任务的多分类问题。

[0072] 步骤3,根据第二概率分布确定金融风险的风险类型。

[0073] 可以将概率值最大的风险类型确定为该金融风险的风险类型。

[0074] 可选地,上述风险主体识别任务的处理流程可以包括以下步骤:

[0075] 步骤1,将待处理文本与金融风险的风险类型进行拼接,并按照预设对应关系转换为第二标记序列。

[0076] 步骤2,将第二标记序列通过第一神经网络模型的嵌入层和编码层,得到第四语义表示向量,第四语义表示向量为包含拼接后的待处理文本和风险类型上下文语义信息的向量。

[0077] 步骤3,对第四语义表示向量进行线性变换,得到第五语义表示向量,第五语义表示向量是通过风险主体识别任务的输出层的私有参数进行处理后得到的。

[0078] 在风险主体识别任务中进行的线性变换是通过风险主体识别任务的输出层的私有参数进行处理后得到的,使得线性变换的结果能够更加适合于风险主体识别。

[0079] 步骤4,利用第五语义表示向量确定第三概率分布和第四概率分布,第三概率分布为第五语义表示向量中的每一个字向量作为与风险类型匹配的风险主体的起始字向量的概率值,第四概率分布为第五语义表示向量中的每一个字向量作为与风险类型匹配的风险主体的终止字向量的概率值。

[0080] 本申请实施例中,每一个字向量都可以对应获得两种概率值,一种是该字向量是与风险类型匹配的风险主体的第一个字向量,即起始字向量的概率,第二种是该字向量是与风险类型匹配的风险主体的最后一个字向量,即终止字向量的概率。

[0081] 步骤5,根据第三概率分布和所述第四概率分布确定与风险类型匹配的风险主体。

[0082] 本申请实施例种,将作为起始字向量概率最高的一个字向量确定为风险主体的第一个字向量,将作为终止字向量概率最高的一个字向量确定为风险主体的最后一个字向量,提取该第一个字向量与该最后一个字向量及其之间的字向量,得到风险主体。

[0083] 本申请技术方案中,考虑到金融风险预测多采用量化金融数据,而量化金融数据有限,而金融文本挖掘模型单一,缺乏对语义的深度挖掘导致模型性能不佳的问题,本申请提供一种基于文本预训练和多任务学习的金融风险预测方法,具体为通过基于预训练语言模型进行多任务学习训练得到的神经网络模型识别待处理文本中是否包含金融风险、风险类型及与风险类型匹配的风险主体的多任务处理,利用预训练语言模型技术解决了缺乏对语义的深度挖掘导致模型性能不佳的问题,并采用多任务处理,解决了数据量有限而且任务之间无法进行信息共享导致模型性能不佳的技术问题。

[0084] 本申请提出了一种基于预训练语言模型进行多任务学习的模型训练方法,如图3所示,进一步详述本申请的技术方案。

[0085] 可选地,步骤S204将待处理文本输入第一神经网络模型之前,该方法还包括按照如下方式对第二神经网络模型进行多任务的训练,得到第一神经网络模型:

[0086] 步骤S302,从训练数据池中随机确定一个批次的训练数据,训练数据包括用于风险识别任务、风险分类任务及风险主体识别任务的训练数据;

[0087] 步骤S304,将训练数据输入第二神经网络模型,并在第二神经网络模型的预训练参数的基础上继续训练第二神经网络模型的各项参数;

[0088] 步骤S306,采用早停的训练方式,在第二神经网络模型对测试数据的识别准确度达到最优值的情况下,将第二神经网络模型作为第一神经网络模型;

[0089] 步骤S308,在第二神经网络模型对测试数据的识别准确度未达到最优值的情况下,继续使用训练数据对第二神经网络模型进行训练,以调整第二神经网络模型内各网络层中的参数的数值,直至第二神经网络模型对测试数据的识别准确度达到最优值。

[0090] 本申请实施例中,模型的训练采取了多任务学习技术中常用的小批次数据训练方法,将训练数据分批次进行多轮训练,在每一个训练轮次中,每次从训练数据池中随机挑选一个批次的训练数据输入模型,计算损失函数,进行反向传播,并更新模型参数,直到所有数据都已被输入,则此轮训练结束。上述早停的训练方式表示将模型训练到在测试集上面表现最优,即停止训练。上述第二神经网络模型为训练状态处于多任务训练前及多任务训练中的模型,上述第一神经网络模型为训练状态在多任务训练完成后的模型。

[0091] 采用本申请的技术方案,利用小批次训练数据对预训练语言模型进行多任务学习的训练,能够使得得到的神经网络模型具有挖掘文本中的深度语义特征的能力,并显著提升每一个任务的识别精确度和速度,提升模型的性能。

[0092] 本申请提出了一种训练金融预训练语言模型的方法,如图4所示,进一步详述本申请的技术方案。

[0093] 可选地,将训练数据输入第二神经网络模型之前,该方法还包括按照如下方式利用无标注的预训练语料对深度神经网络模型进行预训练,得到金融预训练语言模型:

[0094] 步骤S402,获取预训练语料,预训练语料来自于互联网平台金融领域;

[0095] 步骤S404,按照第一预训练语言模型的输入要求对预训练语料进行预处理,第一预训练语言模型为深度神经网络模型,第一预训练语言模型为基于通用领域语料进行预训练得到的预训练语言模型;

[0096] 步骤S406,利用经过预处理的预训练语料对第一预训练语言模型进行预训练;

[0097] 步骤S408,在第一预训练语言模型在目标预训练任务上的性能达到目标性能阈值

的情况下,将第一预训练语言模型作为金融预训练语言模型;

[0098] 步骤S410,在第一预训练语言模型在目标预训练任务上的性能未达到目标性能阈值的情况下,继续使用预训练语料对第一预训练语言模型进行预训练,以调整第一预训练语言模型内各网络层中的参数的数值,直至第一预训练语言模型在目标预训练任务上的性能达到目标性能阈值。

[0099] 预训练语言模型技术是近年来深度学习自然语言处理技术中的研究热点,其核心技术思想是利用大量无标注的预训练语料,对多层深度神经网络模型进行参数初始化,从而使其具有提取到输入文本表示的上下文关联信息的能力。在处理下游具体自然语言处理任务的时候,仅需要对预训练好的大部分模型参数进行微调,并依据具体任务目标在模型下游增加对应的输出层,即可得到比较理想的任务效果。

[0100] 本申请实施例中,预训练语料可以是互联网上(具体可以是各大包含金融新闻的门户网站,如新浪网、同花顺网、云财经网等)爬取的足量金融新闻标题与正文。

[0101] 本申请实施例中,第一预训练语言模型可以是2018年Google发布的BERT模型,还可以是其它预训练语言模型。以BERT为代表的通用预训练语言模型均为基于通用领域语料进行预训练得到。而诸如金融此类的特定领域语料,在文本特征和语法结构上有一定的特殊性和规律性,譬如,常有金融领域专业名词出现(如减持、停牌、披露、套现等),并在语句结构上一般呈现书面化、长难句较多的特点。针对以上问题,可以基于BERT模型,利用从互联网爬取的无标注金融语料对模型进行进一步的预训练。

[0102] 上述按照第一预训练语言模型的输入要求对预训练语料进行预处理可以包括格式调整、令牌化(tokenize,即转化为机器更容易处理的标记符号)与词遮蔽等。

[0103] 本申请实施例中,利用预训练语言模型在两个预训练任务MLM(遮蔽语言模型)和NSP(下一语句预测)上的性能表现,来控制训练的迭代数和轮数。

[0104] 采用本申请技术方案,能够得到在金融领域任务上具有更优异性能表现得金融领域预训练语言模型。

[0105] 可选地,将训练数据输入第二神经网络模型之前,该方法还包括按照如下方式结合金融预训练语言模型,得到该第二神经网络模型:

[0106] 在金融预训练语言模型的输出层分别添加用于风险识别任务、风险分类任务及风险主体识别任务的输出层,得到第二神经网络模型。

[0107] 本申请实施例中,对金融预训练语言模型增加用于多任务处理的输出层,得到多任务学习模型,即该第二神经网络模型,以进行多任务学习及预测。

[0108] 可选地,如图5所示,从训练数据池中随机确定一个批次的训练数据之前,该方法还包括按照如下方式构建训练数据池:

[0109] 步骤S502,按照预设的每个批次的的数据量大小将用于风险识别任务、风险分类任务及风险主体识别任务的训练数据划分为多个批次;

[0110] 步骤S504,将所有批次的训练数据进行无序混合,得到训练数据池。

[0111] 本申请实施例中,构造训练数据池以应用于模型训练时小批次数据训练方法。在数据集构造方面,本申请实施例构造了两个粒度不同的金融风险文本数据集,用于实验和验证模型效果(下文中称为数据集一和数据集二)。两个数据集的数据来源分别来自于蚂蚁金服内部的咨询系统(后被发布于CCKS2018评测任务四中)和微博微信爬取,规模分别为24

万余条和15万余条,都被按照模型输入要求对数据格式进行了处理,标注了有无风险,并对有风险的数据样本标注了风险类型、与之对应的风险主体。

[0112] 可选地,如图6所示,进行多任务学习的模型训练过程中,在第二神经网络模型的预训练参数的基础上继续训练第二神经网络模型的各项参数包括:

[0113] 步骤S602,将第二神经网络模型的嵌入层和编码层作为共享参数区,将第二神经网络模型的各个输出层分别作为私有参数区,私有参数区包括第一私有参数区、第二私有参数区及第三私有参数区,第一私有参数区为风险识别任务的输出层,第二私有参数区为风险分类任务的输出层,第三私有参数区为风险主体识别任务的输出层;

[0114] 步骤S604,将第一私有参数区、第二私有参数区及第三私有参数区的学习率固定为第一学习率,并利用训练数据对第二神经网络模型进行训练,以在多个第二学习率中确定共享参数区的第一目标学习率,第一目标学习率为适用于共享参数区的最优学习率;

[0115] 步骤S606,将共享参数区的学习率固定为第一目标学习率,并利用训练数据对第二神经网络模型进行训练,以在目标范围内分别确定第一私有参数区、第二私有参数区及第三私有参数区的第二目标学习率,第二目标学习率为分别适用于第一私有参数区、第二私有参数区及第三私有参数区的最优学习率。

[0116] 本申请实施例中,提供了一个能完成三个任务目标的多任务学习架构,其多任务学习的特性依靠模型部分参数的共享来实现,模型的嵌入层和编码层部分将被共享,后续三个任务会具有各自的输出层,以区分不同任务。

[0117] 考虑到风险识别任务、风险分类任务及风险主体识别任务在模型训练中可能具有不同的收敛速度,以及共享层在学习文本表示时具有鲁棒性,因此本申请实施例对模型共享部分参数(即嵌入层和编码层)、三个任务的私有部分参数(即三个任务各自输出层)均设置了不同的学习率。在分别确定模型各部分学习率的过程中,本申请实施例首先可以固定私有部分参数学习率为 $1e-3$,然后可以根据BERT模型设置 $2e-5$ 、 $3e-5$ 和 $5e-5$ 三种共享部分的学习率,探究最优学习率;然后固定共享部分的学习率为探究到的最有学习率,将私有部分参数设置为 0.1 、 0.01 …直至 $1e-6$,目的是探究私有部分参数的尺度;确定私有部分参数尺度之后,按照每次乘以二的方法进行私有部分参数学习率的细调,确定最终的私有部分最佳学习率。经实验表明,此种设置下的模型性能显著优于全模型采取相同学习率设置下的模型。

[0118] 可选地,进行多任务学习的模型训练过程中,在第二神经网络模型的预训练参数的基础上继续训练第二神经网络模型的各项参数还包括:

[0119] 通过共享参数区的参数共享,在训练第二神经网络模型的过程中确定目标隐层参数,目标隐层参数为同时适用于第一私有参数区、第二私有参数区及第三私有参数区的隐层参数。

[0120] 本申请实施例中,可以通过共享编码层参数,使得模型过拟合的风险被大幅降低,模型在训练过程中将努力学习到一个适合全部三个任务的隐层表征。

[0121] 多任务学习(Multi-task Learning)受到人类学习活动的启发,人们通常将从先前任务中学到的知识应用于帮助学习新任务。譬如说,一个会滑雪的人学习溜冰往往会比一个不会滑雪的人更加容易。因此采用本申请技术方案,在风险识别、风险分类和风险主体识别三个任务之间进行信息互通和共享,显著提升模型的性能。

[0122] 例如,金融文本为“投资返利?还卖原始股?记者暗访‘重庆聚才道集团’可能存在非法集资活动嫌疑”。显然,句中的“非法集资活动”这几个汉字是风险识别任务把这条样本预测为“有风险”的主要依据,这几个汉字所对应的令牌(Token)中蕴含的信息在最后的隐层表示中也为主要的特征;在后续的风险分类任务中,这几个字也同样是模型把将这条样本分类为“涉嫌非法集资”类别的主要判断依据。最后的风险主体识别任务中,模型也可以首先定位到“存在非法集资”这几个字的附近,然后搜寻相应的事件主体,才可正确抽出“聚才道”为识别结果。因此,任务之间进行信息互通和共享对各个任务都具有积极意义。

[0123] 而在本申请实施例中,对训练数据进行标记及模型输出标签可以是:训练数据在输入模型前会经历预处理,文本被令牌化,利用标记信息标记标签,即对标签进行编号(例如,风险识别中,有风险和无风险分别编号为“0”和“1”;风险分类中,若共有16种风险类别,将会被一一对应成“0”-“15”的标签)。在输出层,对于风险识别任务,模型输出的将会是标签“0”或标签“1”;对于风险分类任务,模型输出的是模型预测金融风险类别所对应的标签编号;对于风险主体识别任务,模型输出的将是模型预测答案在输入文本中的开始位置与结束位置(例如,若输入文本为“天龙光电(300029)或成创业板首家退市公司谭力被查后海韵集团董事长曾被采取强制措施”,模型预测答案为“海韵集团”,则模型输出结果会是“28”和“31”,因28和31是“海韵集团”在原输入文本中的开始位置和结束位置)。

[0124] 本申请提出了一种串行模型测试方法,如图7所示,进一步详述本申请的技术方案。

[0125] 可选地,在第二神经网络模型对测试数据的识别准确度达到最优值的情况下,将第二神经网络模型作为第一神经网络模型可以包括以下步骤:

[0126] 步骤S702,获取第一测试数据。

[0127] 步骤S704,将第一测试数据输入第二神经网络模型,以按照风险识别任务的处理流程对第一测试数据进行处理,得到风险识别任务的输出层输出的风险识别结果。

[0128] 步骤S706,确定风险识别结果的精确率与召回率的第一调和平均值,并筛选出第二测试数据,第二测试数据为标记信息为有风险且风险识别结果为有风险的第一测试数据。

[0129] 步骤S708,利用第二神经网络模型按照风险分类任务的处理流程对第二测试数据进行处理,得到风险分类任务的输出层输出的风险分类结果。

[0130] 步骤S710,确定风险分类结果的准确率和风险分类结果的排序倒数值,并筛选出第三测试数据,第三测试数据为风险分类结果与标记信息标记的风险类型匹配的第二测试数据。

[0131] 步骤S712,利用第二神经网络模型按照风险主体识别任务的处理流程对第三测试数据进行处理,得到风险主体识别任务的输出层输出的风险主体识别结果。

[0132] 步骤S714,确定风险主体识别结果的完全匹配值和风险主体识别结果的精确率与召回率的第二调和平均值。

[0133] 步骤S716,在第一调和平均值、准确率、排序倒数值、完全匹配值以及第二调和平均值达到对应的预设指标的情况下,将第二神经网络模型确定为所述第一神经网络模型。

[0134] 需要说明的是,在对第二神经网络模型进行训练,还未训练完成得到第一神经网络模型的过程中,上述风险识别任务的处理流程、风险分类任务的处理流程及风险主体识

别任务的处理流程都是基于还未完成训练的多任务学习模型即该第二神经网络模型进行的。

[0135] 本申请实施例中,可以依照任务特性与常见标准为三个任务设计不同的评价指标。

[0136] 风险识别即识别输入文本是否包含金融风险,是一个文本二分类任务,且在实验数据和实际应用中具有正负样本数量不平均的特点,负样本(“无金融风险”)的数据样本一般数量为正样本(“有金融风险”样本)的几倍之多。参考常用的对具有正负例不平均特点的二分类问题的评价指标,本申请可以采取F1值,即精确率与召回率的调和平均值作为风险识别任务的模型性能的评价指标。

[0137] 风险分类即对前一步已经判断为“有风险”的数据样本进行风险类型的定性分类,是一个文本多分类任务,本申请实施例可以采用分类的ACC值(即Accuracy,准确率)和MRR值(即Mean reciprocal rank,排序倒数)作为风险分类任务的模型性能的评价指标。

[0138] 风险主体识别,即对前一步中已经判定金融风险类型的数据,进一步在文本中根据其风险类型抽取对应的风险主体,是一个涉及消歧的命名实体识别任务。本申请实施例中可以使用机器阅读理解模型以进行金融风险主体识别,解决歧义问题。因此参考对机器阅读理解模型常见的评价指标,可以采用EM值(Exact match,完全匹配)和F1值来作为风险主体识别任务的模型性能的评价指标。

[0139] 本申请在通用领域预训练语言模型已有参数基础上,使用互联网爬取的金融文本语料进行了继续预训练,旨在更细粒度地捕获金融领域文本数据的语义特征。同时,本申请采用多任务学习技术思想设计模型结构,将三个任务对应模型的嵌入层和编码层进行参数共享,以实现不同任务之间信息的流通,解决了模型实际应用中可能出现的训练数据量不够充分的问题,进一步利用风险识别、风险分类和风险主体识别三个任务之间的可共享信息。

[0140] 本申请实施例建立在大量实验的基础上,经实验验证表明,本申请所得到的预训练语言模型(下称为“金融预训练语言模型”)在金融领域自然语言处理任务上有更优秀的性能表现:在上述的数据集一上,保持其他实验条件相同,金融预训练语言模型能在风险识别任务上取得0.5%的F1值性能提高(93.1%~93.6%);在风险分类任务中,ACC指标提升了0.9%(90.0%~90.9%),MRR指标提升了1.2%(91.0%~92.2%);在风险主体识别任务中,F1指标提高了2.4%(71.3%~73.7%),EM指标提升了2.7%(58.4%~61.1%)。

[0141] 实验表明,相较于为分别为各任务单独设置模型,多任务学习模型架构设计对每个任务的性能基本都有提升,对模型整体性能有明显提升。在数据集一上,保持其他实验条件相同,多任务学习模型相较于单模型,在风险识别任务上性能略微降低(93.1%~93.0%);在风险分类任务中,ACC指标提升了0.6%(90.0%~90.6%),MRR指标提升了1.2%(91.0%~91.9%);在风险主体识别任务中,F1指标提高了0.1%(71.3%~71.4%),EM指标提升了1.5%(58.4%~59.9%)。在数据集二上,保持其他实验条件相同,多任务学习模型相较于单模型,在风险识别任务上性能略微降低(99.9%~99.7%);在风险分类任务中,ACC指标提升了0.6%(88.2%~88.8%),MRR指标提升了0.2%(93.1%~93.3%);在风险主体识别任务中,F1指标提高了1.8%(74.1%~75.9%),EM指标提升了0.7%(54.6%~55.3%)。

[0142] 本申请实施例中,不同于传统的命名实体识别任务中常用的序列标注方法,由于还牵涉到与风险类型的对应关系,本申请实施例采用机器阅读理解的方式在抽取主体时进行消歧。例如,待处理文本为“西安饮食:控股股东股份转让,受让主体变更为华侨城文化集团”,此文本中包含的金融风险类型为“实控人股东变更”,而该文本中存在的实体有“西安饮食”和“华侨城文化集团”,而与“实控人股东变更”所对应的风险事件主体应只为“华侨城文化集团”,而不是“西安饮食”。相对于传统的命名实体识别问题来说,本申请采用机器阅读理解的方式具有分辨提取出的命名实体是否正确对应风险类型的能力,即能够消歧,从而使识别结果更加准确。

[0143] 本申请利用预训练语言模型技术,解决了缺乏对语义的深度挖掘导致模型性能不佳的问题,并采用多任务处理,解决了数据量有限而且任务之间无法进行信息共享导致模型性能不佳的技术问题。此外,多任务处理的方法降低了模型参数总量,具有节约存储空间并提升模型加载运行速度的效果。

[0144] 根据本申请实施例的又一方面,如图8所示,提供了一种基于文本预训练和多任务学习的金融风险预测装置,包括:获取模块801,用于获取待处理文本,待处理文本来自于互联网平台金融领域;风险识别模块803,用于将待处理文本输入第一神经网络模型,以按照风险识别任务的处理流程确定待处理文本的内容是否包括金融风险,第一神经网络模型是采用具有标记信息的训练数据对第二神经网络模型进行多任务的训练得到的,第二神经网络模型为结合金融预训练语言模型的多任务学习模型,金融预训练语言模型为利用多个无标注的预训练语料进行参数初始化后得到的预训练语言模型,多任务包括风险识别任务、风险分类任务及风险主体识别任务,标记信息用于标记训练数据的内容是否包括金融风险,在包括金融风险的情况下还用于标记金融风险的风险类型,并标记与风险类型匹配的风险主体;风险分类模块805,用于在待处理文本的内容包括金融风险的情况下,利用第一神经网络模型按照风险分类任务的处理流程确定金融风险的风险类型;风险主体识别模块807,用于利用第一神经网络模型按照风险主体识别任务的处理流程确定与风险类型匹配的风险主体。

[0145] 需要说明的是,该实施例中的获取模块801可以用于执行本申请实施例中的步骤S202,该实施例中的风险识别模块803可以用于执行本申请实施例中的步骤S204,该实施例中的风险分类模块805可以用于执行本申请实施例中的步骤S206,该实施例中的风险主体识别模块807可以用于执行本申请实施例中的步骤S208。

[0146] 此处需要说明的是,上述模块与对应的步骤所实现的示例和应用场景相同,但不限于上述实施例所公开的内容。需要说明的是,上述模块作为装置的一部分可以运行在如图1所示的硬件环境中,可以通过软件实现,也可以通过硬件实现。

[0147] 可选地,该基于文本预训练和多任务学习的金融风险预测装置,还包括:训练数据提取模块,用于从训练数据池中随机确定一个批次的训练数据,训练数据包括用于风险识别任务、风险分类任务及风险主体识别任务的训练数据;第一训练模块,用于将训练数据输入第二神经网络模型,并在第二神经网络模型的预训练参数的基础上继续训练第二神经网络模型的各项参数;第二训练模块,用于采用早停的训练方式,在第二神经网络模型对测试数据的识别准确度达到最优值的情况下,将第二神经网络模型作为第一神经网络模型;第三训练模块,用于在第二神经网络模型对测试数据的识别准确度未达到最优值的情况下,

继续使用训练数据对第二神经网络模型进行训练,以调整第二神经网络模型内各网络层中的参数的数值,直至第二神经网络模型对测试数据的识别准确度达到最优值。

[0148] 可选地,该基于文本预训练和多任务学习的金融风险预测装置,还包括:预训练语料获取模块,用于获取预训练语料,预训练语料来自于互联网平台金融领域;预处理模块,用于按照第一预训练语言模型的输入要求对预训练语料进行预处理,第一预训练语言模型为深度神经网络模型,第一预训练语言模型为基于通用领域语料进行预训练得到的预训练语言模型;第一预训练模块,用于利用经过预处理的预训练语料对第一预训练语言模型进行预训练;第二预训练模块,用于在第一预训练语言模型在目标预训练任务上的性能达到目标性能阈值的情况下,将第一预训练语言模型作为金融预训练语言模型;第三预训练模块,用于在第一预训练语言模型在目标预训练任务上的性能未达到目标性能阈值的情况下,继续使用预训练语料对第一预训练语言模型进行预训练,以调整第一预训练语言模型内各网络层中的参数的数值,直至第一预训练语言模型在目标预训练任务上的性能达到目标性能阈值。

[0149] 可选地,该基于文本预训练和多任务学习的金融风险预测装置,还包括:多任务处理模块,用于在金融预训练语言模型的输出层分别添加用于风险识别任务、风险分类任务及风险主体识别任务的输出层,得到第二神经网络模型。

[0150] 可选地,该基于文本预训练和多任务学习的金融风险预测装置,还包括:批次划分模块,用于按照预设的每个批次的的数据量大小将用于风险识别任务、风险分类任务及风险主体识别任务的训练数据划分为多个批次;混合模块,用于将所有批次的训练数据进行无序混合,得到训练数据池。

[0151] 可选地,该基于文本预训练和多任务学习的金融风险预测装置,还包括:参数区确定模块,用于将第二神经网络模型的嵌入层和编码层作为共享参数区,将第二神经网络模型的各个输出层分别作为私有参数区,私有参数区包括第一私有参数区、第二私有参数区及第三私有参数区,第一私有参数区为风险识别任务的输出层,第二私有参数区为风险分类任务的输出层,第三私有参数区为风险主体识别任务的输出层;共享参数区学习率确定模块,用于将第一私有参数区、第二私有参数区及第三私有参数区的学习率固定为第一学习率,并利用训练数据对第二神经网络模型进行训练,以在多个第二学习率中确定共享参数区的第一目标学习率,第一目标学习率为适用于共享参数区的最优学习率;私有参数区学习率确定模块,用于将共享参数区的学习率固定为第一目标学习率,并利用训练数据对第二神经网络模型进行训练,以在目标范围内分别确定第一私有参数区、第二私有参数区及第三私有参数区的第二目标学习率,第二目标学习率为分别适用于第一私有参数区、第二私有参数区及第三私有参数区的最优学习率。

[0152] 可选地,该基于文本预训练和多任务学习的金融风险预测装置,还包括:参数共享模块,用于通过共享参数区的参数共享,在训练第二神经网络模型的过程中确定目标隐层参数,目标隐层参数为同时适用于第一私有参数区、第二私有参数区及第三私有参数区的隐层参数。

[0153] 可选地,该第二训练模块还用于:获取第一测试数据;将第一测试数据输入第二神经网络模型,以按照风险识别任务的处理流程对第一测试数据进行处理,得到风险识别任务的输出层输出的风险识别结果;确定风险识别结果的精确率与召回率的第一调和平均

值,并筛选出第二测试数据,第二测试数据为标记信息为有风险且风险识别结果为有风险的第一测试数据;利用第二神经网络模型按照风险分类任务的流程对第二测试数据进行处理,得到风险分类任务的输出层输出的风险分类结果;确定风险分类结果的准确率和风险分类结果的排序倒数值,并筛选出第三测试数据,第三测试数据为风险分类结果与标记信息标记的风险类型匹配的第二测试数据;利用第二神经网络模型按照风险主体识别任务的流程对第三测试数据进行处理,得到风险主体识别任务的输出层输出的风险主体识别结果;确定风险主体识别结果的完全匹配值和风险主体识别结果的精确率与召回率的第二调和平均值;在第一调和平均值、准确率、排序倒数值、完全匹配值以及第二调和平均值达到对应的预设指标的情况下,将第二神经网络模型确定为所述第一神经网络模型。

[0154] 可选地,该风险识别模块还用于:按照预设对应关系将待处理文本转换为第一标记序列;将第一标记序列通过第一神经网络模型的嵌入层和编码层,得到所述待处理文本的第一语义表示向量,第一语义表示向量为包含待处理文本上下文语义信息的向量;对第一语义表示向量进行线性变换,得到第二语义表示向量,第二语义表示向量是通过风险识别任务的输出层的私有参数进行处理后得到的;采用Softmax分类的方式对第二语义表示向量进行处理,得到第一概率分布,第一概率分布为待处理文本的内容包含金融风险的概率值,第一概率分布是通过风险识别任务的输出层得到的;根据第一概率分布确定待处理文本是否包含金融风险。

[0155] 可选地,该风险分类模块还用于:对第一语义表示向量进行线性变换,得到第三语义表示向量,第三语义表示向量是通过风险分类任务的输出层的私有参数进行处理后得到的;采用Softmax分类的方式对第三语义表示向量进行处理,得到第二概率分布,第二概率分布为金融风险的风险类型属于各个类型的概率值,第二概率分布是通过风险分类任务的输出层得到的;根据第二概率分布确定金融风险的风险类型。

[0156] 可选地,该风险主体识别模块还用于:将待处理文本与金融风险的风险类型进行拼接,并按照预设对应关系转换为第二标记序列;将第二标记序列通过第一神经网络模型的嵌入层和编码层,得到第四语义表示向量,第四语义表示向量为包含拼接后的待处理文本和风险类型上下文语义信息的向量;对第四语义表示向量进行线性变换,得到第五语义表示向量,第五语义表示向量是通过风险主体识别任务的输出层的私有参数进行处理后得到的;利用第五语义表示向量确定第三概率分布和第四概率分布,第三概率分布为第五语义表示向量中的每一个字向量作为与风险类型匹配的风险主体的起始字向量的概率值,第四概率分布为第五语义表示向量中的每一个字向量作为与风险类型匹配的风险主体的终止字向量的概率值;根据第三概率分布和所述第四概率分布确定与风险类型匹配的风险主体。

[0157] 根据本申请实施例的另一方面还提供了一种计算机设备,包括存储器、处理器,所述存储器中存储有可在所述处理器上运行的计算机程序,所述处理器执行所述计算机程序时实现上述步骤。

[0158] 上述计算机设备中的存储器、处理器通过通信总线和通信接口进行通信。所述通信总线可以是外设部件互连标准(Peripheral Component Interconnect,简称PCI)总线或扩展工业标准结构(Extended Industry Standard Architecture,简称EISA)总线等。该通信总线可以分为地址总线、数据总线、控制总线等。

[0159] 存储器可以包括随机存取存储器(Random Access Memory,简称RAM),也可以包括非易失性存储器(non-volatile memory),例如至少一个磁盘存储器。可选的,存储器还可以是至少一个位于远离前述处理器的存储装置。

[0160] 上述的处理器可以是通用处理器,包括中央处理器(Central Processing Unit,简称CPU)、网络处理器(Network Processor,简称NP)等;还可以是数字信号处理器(Digital Signal Processing,简称DSP)、专用集成电路(Application Specific Integrated Circuit,简称ASIC)、现场可编程门阵列(Field-Programmable Gate Array,简称FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。

[0161] 根据本申请实施例的又一方面还提供了一种具有处理器可执行的非易失的程序代码的计算机可读介质。

[0162] 可选地,在本申请实施例中,计算机可读介质被设置为存储用于所述处理器执行以下步骤的程序代码:

[0163] 步骤S202,获取待处理文本,待处理文本来自于互联网平台金融领域;

[0164] 步骤S204,将待处理文本输入第一神经网络模型,以按照风险识别任务的处理流程确定待处理文本的内容是否包括金融风险,第一神经网络模型是采用具有标记信息的训练数据对第二神经网络模型进行多任务的训练得到的,第二神经网络模型为结合金融预训练语言模型的多任务学习模型,金融预训练语言模型为利用多个无标注的预训练语料进行参数初始化后得到的预训练语言模型,多任务包括风险识别任务、风险分类任务及风险主体识别任务,标记信息用于标记训练数据的内容是否包括金融风险,在包括金融风险的情况下还用于标记金融风险的风险类型,并标记与风险类型匹配的风险主体;

[0165] 步骤S206,在待处理文本的内容包括金融风险的情况下,利用第一神经网络模型按照风险分类任务的处理流程确定金融风险的风险类型;

[0166] 步骤S208,利用第一神经网络模型按照风险主体识别任务的处理流程确定与风险类型匹配的风险主体。

[0167] 可选地,本实施例中的具体示例可以参考上述实施例中所描述的示例,本实施例在此不再赘述。

[0168] 本申请实施例在具体实现时,可以参阅上述各个实施例,具有相应的技术效果。

[0169] 可以理解的是,本文描述的这些实施例可以用硬件、软件、固件、中间件、微码或其组合来实现。对于硬件实现,处理单元可以实现在一个或多个专用集成电路(Application Specific Integrated Circuits,ASIC)、数字信号处理器(Digital Signal Processing,DSP)、数字信号处理设备(DSP Device,DSPD)、可编程逻辑设备(Programmable Logic Device,PLD)、现场可编程门阵列(Field-Programmable Gate Array,FPGA)、通用处理器、控制器、微控制器、微处理器、用于执行本申请所述功能的其它电子单元或其组合中。

[0170] 对于软件实现,可通过执行本文所述功能的单元来实现本文所述的技术。软件代码可存储在存储器中并通过处理器执行。存储器可以在处理器中或在处理器外部实现。

[0171] 本领域普通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,取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能,但是这种实现不应认为超出

本申请的范围。

[0172] 所属领域的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。

[0173] 在本申请所提供的实施例中,应该理解到,所揭露的装置和方法,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,所述模块的划分,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,例如多个模块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。另一点,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。

[0174] 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方案的目的。

[0175] 另外,在本申请各个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。

[0176] 所述功能如果以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。基于这样的理解,本申请实施例的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的部分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,包括若干指令用以使得一台计算机设备(可以是个人计算机,服务器,或者网络设备等)执行本申请各个实施例所述方法的全部或部分步骤。而前述的存储介质包括:U盘、移动硬盘、ROM、RAM、磁碟或者光盘等各种可以存储程序代码的介质。需要说明的是,在本文中,诸如“第一”和“第二”等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来,而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者设备所固有的要素。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、物品或者设备中还存在另外的相同要素。

[0177] 以上所述仅是本申请的具体实施方式,使本领域技术人员能够理解或实现本申请。对这些实施例的多种修改对本领域的技术人员来说将是显而易见的,本文中所定义的一般原理可以在不脱离本申请的精神或范围的情况下,在其它实施例中实现。因此,本申请将不会被限制于本文所示的这些实施例,而是要符合与本文所申请的原理和新颖特点相一致的最宽的范围。

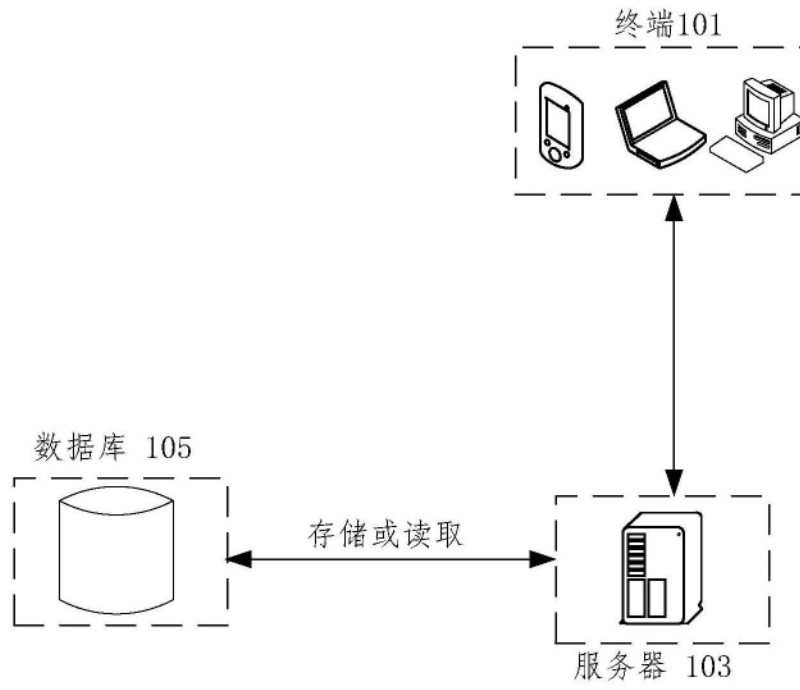


图1

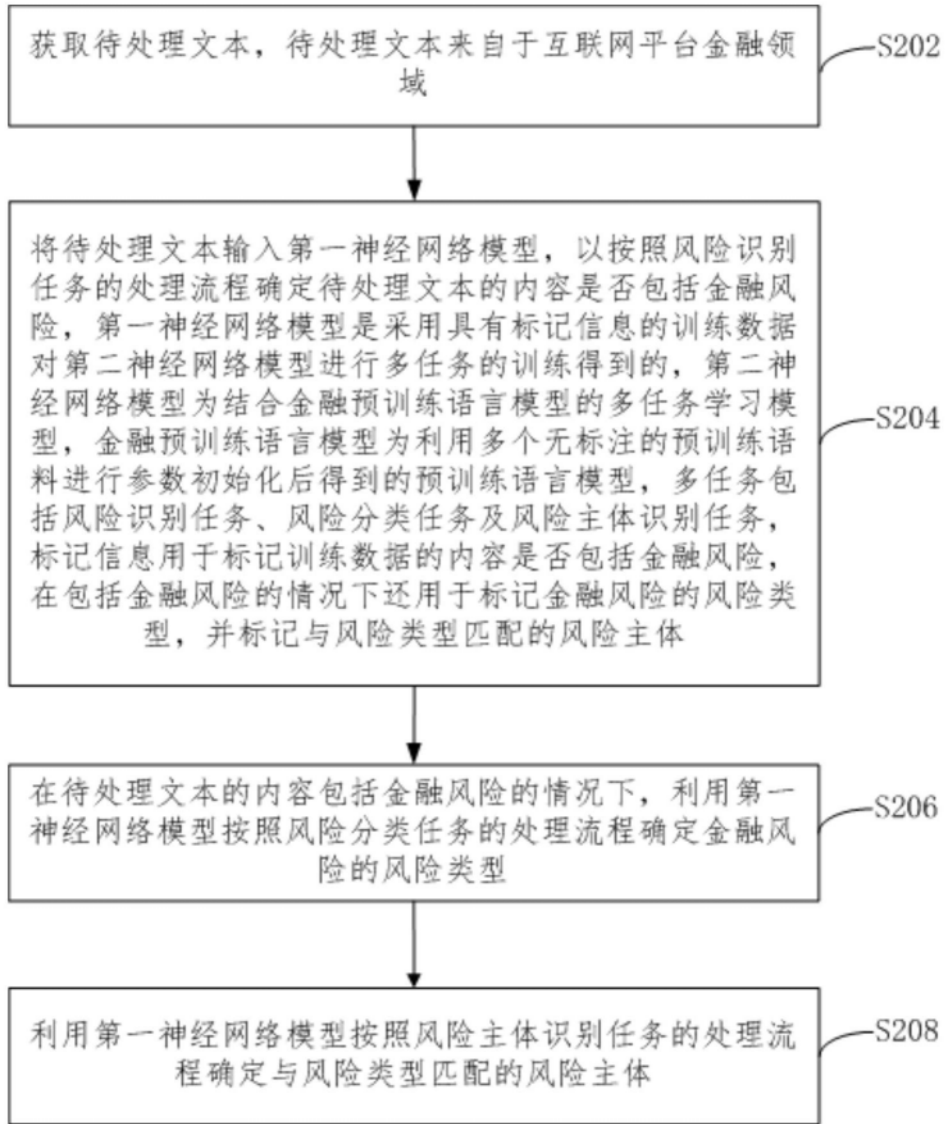


图2

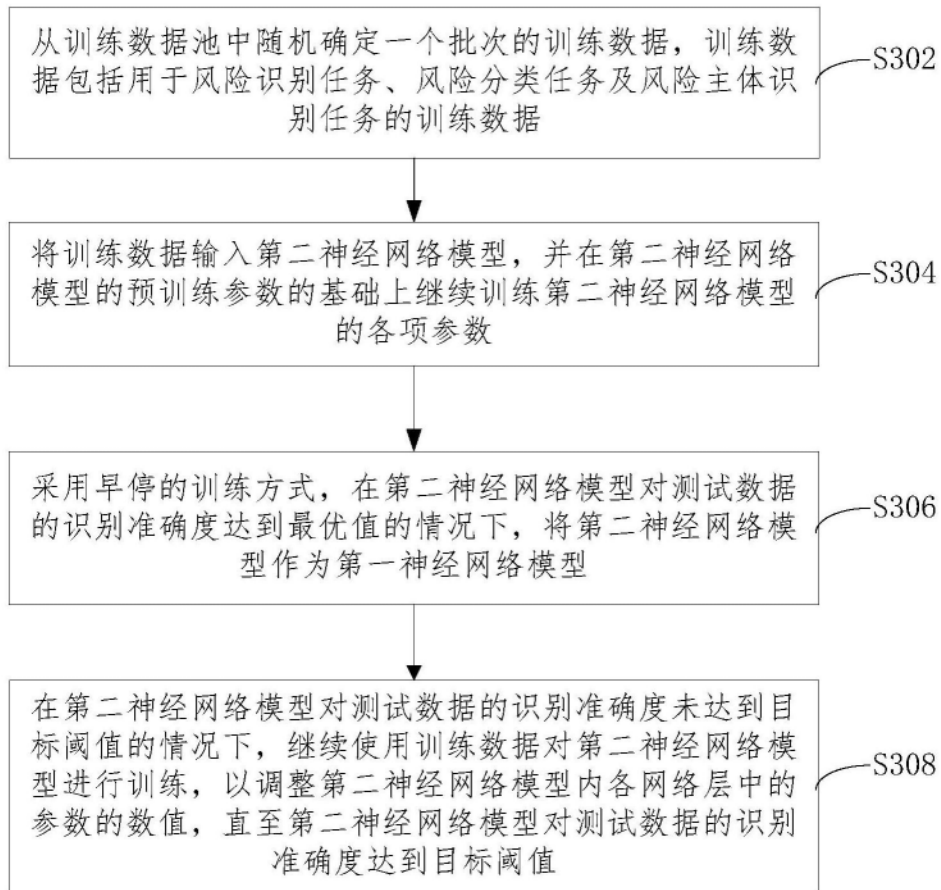


图3

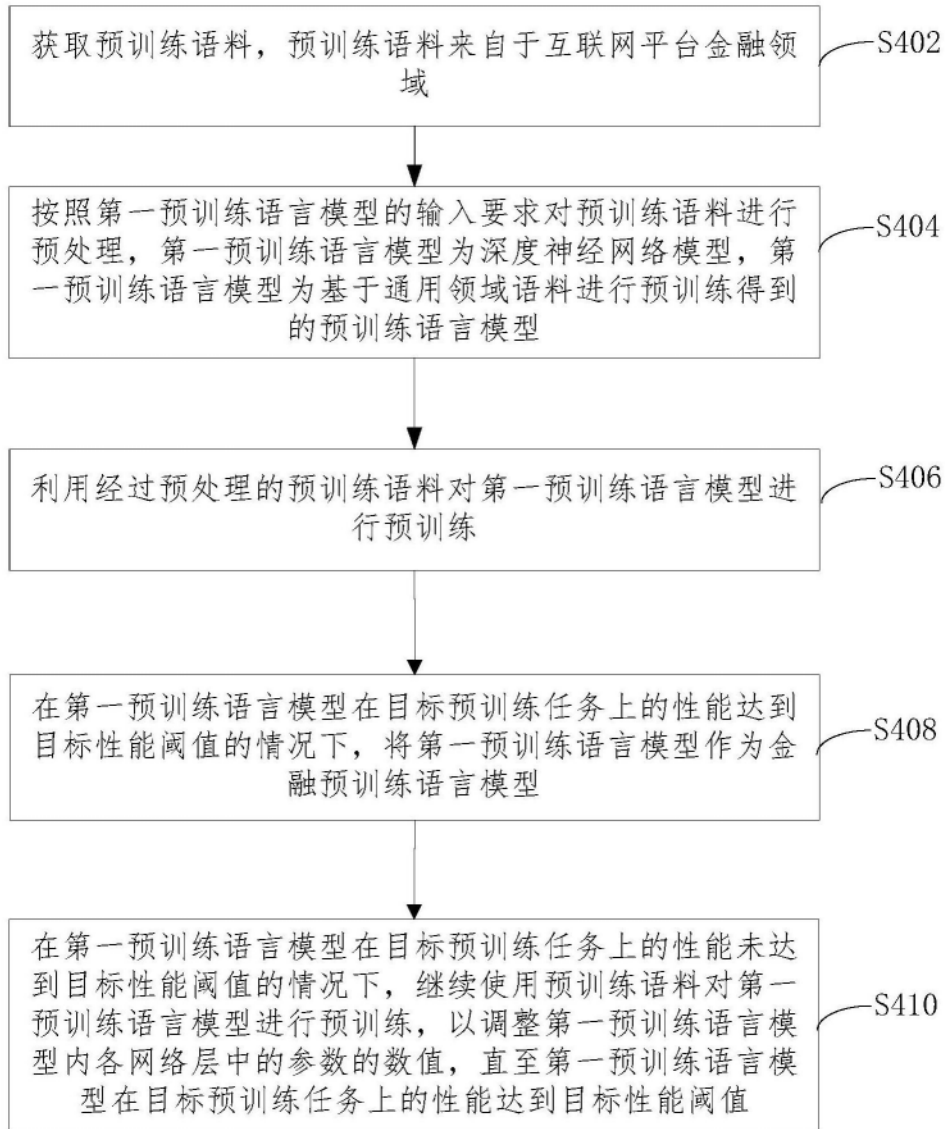


图4

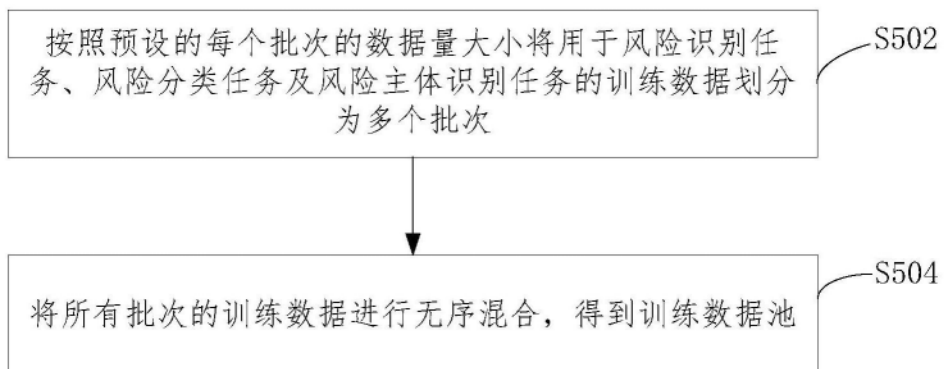


图5

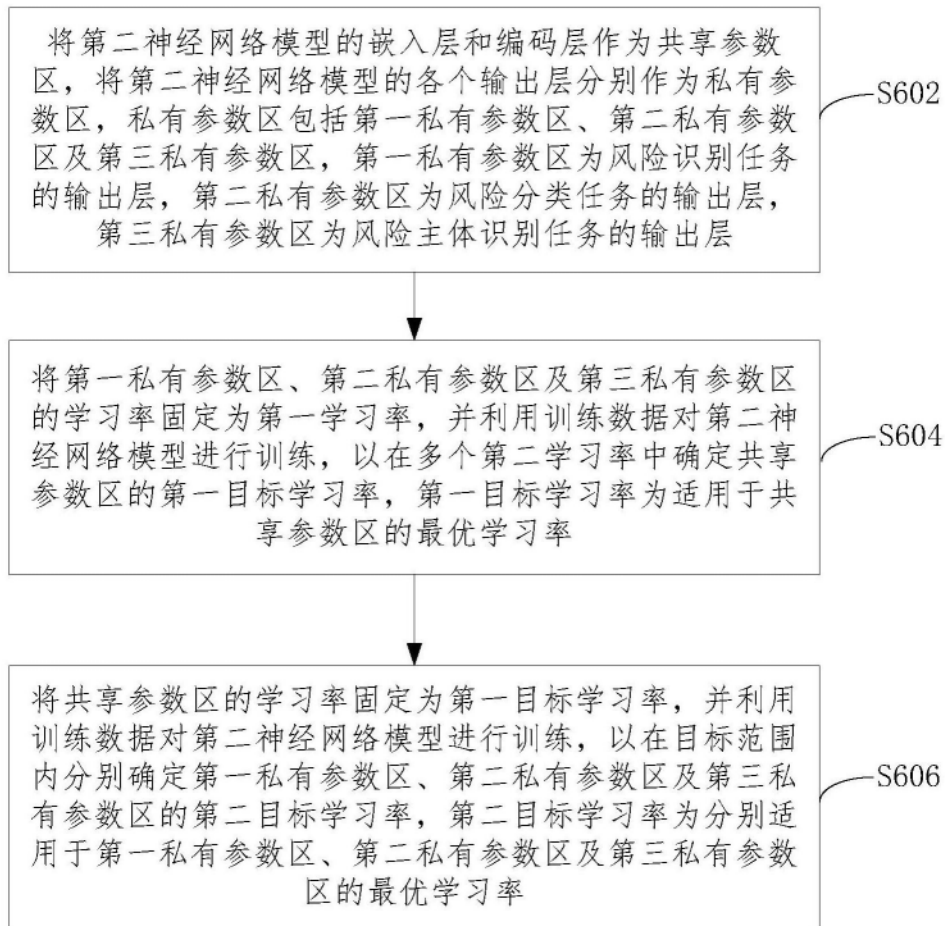


图6

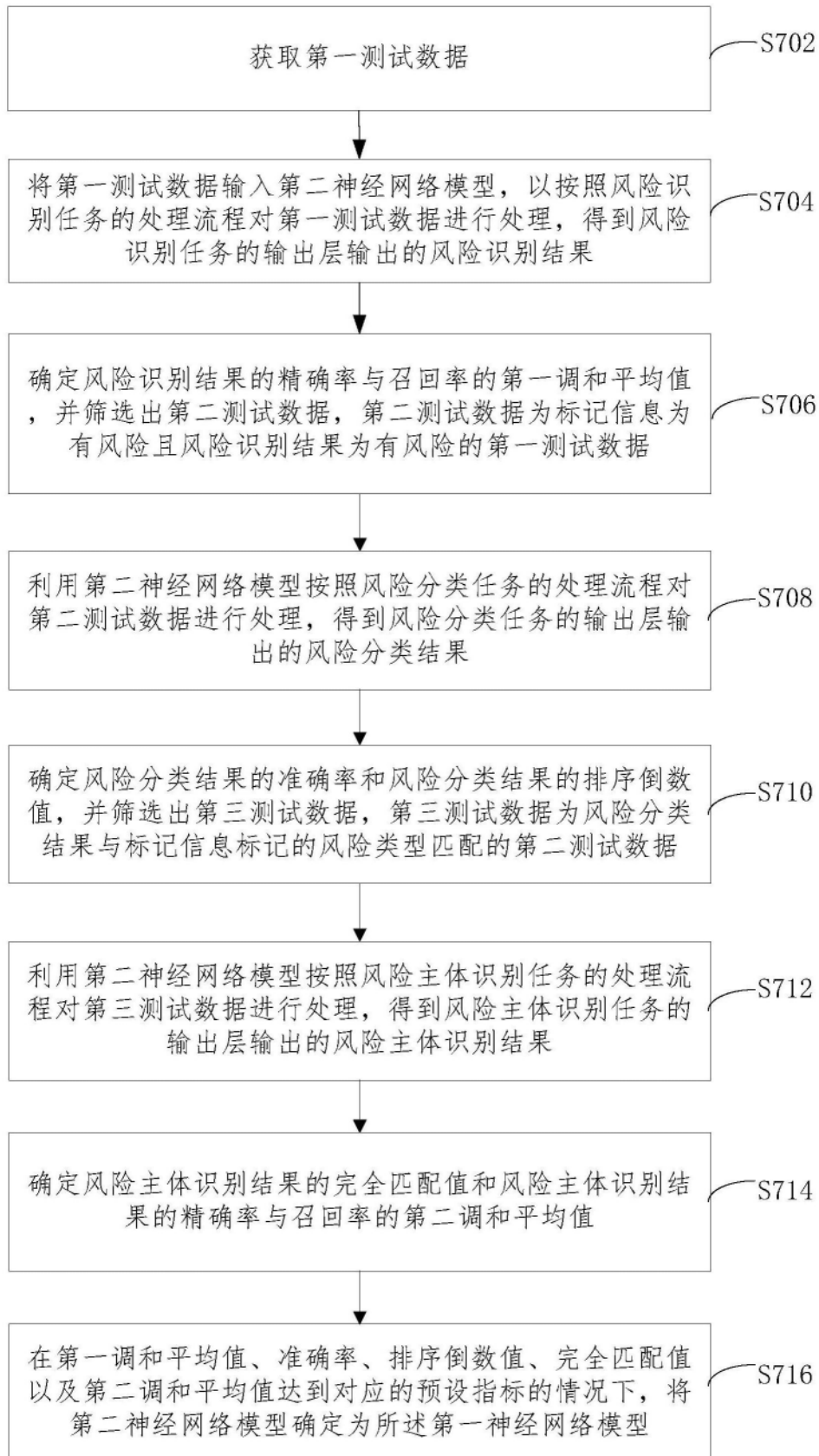


图7

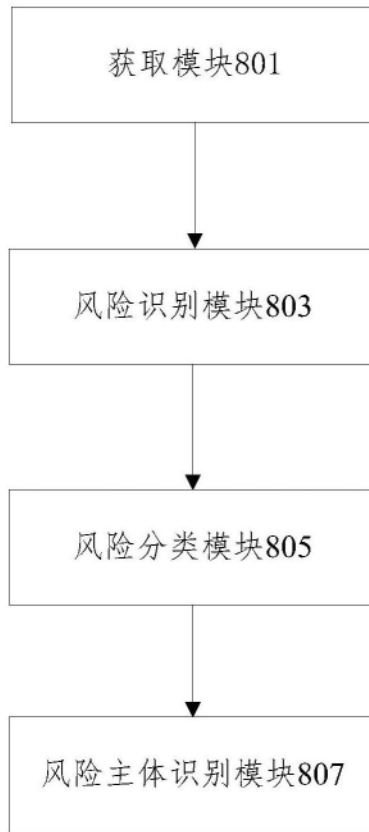


图8