

数据处理方法、装置及设备

技术领域

[01]本文件涉及数据处理技术领域，尤其涉及一种数据处理方法、装置及设备。

背景技术

[02]随着计算机技术的快速发展，网络服务市场的规模日趋庞大，但网络服务的不断发展，也为恶意第三方提供了新的平台，恶意第三方可以通过具有隐蔽含义的暗语绕过风控防控系统进行违法活动，由于具有隐蔽含义的暗语通常与无风险的词语的相似度较高，仅通过词语匹配的方式无法准确的识别出暗语。

[03]可以通过人工基于暗语的上下文，判断当前场景是否存在风险，但待识别的对象的数据量较大，通过人工判断的方式的数据处理效率低、数据处理准确性差，这就导致风险防控的效率和准确率低，因此，需要一种能够在风控场景下，提高针对暗语的风险防控效率和准确率的解决方案。

发明内容

[04]本说明书实施例的目的是提供一种能够在风控场景下，提高针对暗语的风险防控效率和准确率的解决方案。

[05]为了实现上述技术方案，本说明书实施例是这样实现的：第一方面，本说明书实施例提供了一种数据处理方法，包括：获取待识别的目标对象；若所述目标对象中包含与第一暗语匹配的词语，则从预先构建的语料库包含的语料中，获取与所述目标对象对应的目标语料，所述预先构建的语料库包括第一语料，所述第一语料为基于第二暗语和目标风险语料构建的风险语料，所述目标风险语料包含与所述第二暗语具有预设关联关系的风险词语；基于所述目标对象和所述目标语料的相似度，以及所述目标语料的风险标签，确定所述目标对象是否为存在风险。

[06]第二方面，本说明书实施例提供了一种数据处理装置，包括：对象获取模块，用于获取待识别的目标对象；语料获取模块，用于若所述目标对象中包含与第一暗语匹配的词语，则从预先构建的语料库包含的语料中，获取与所述目标对象对应的目标语料，所述预先构建的语料库包括第一语料，所述第一语料为基于第二暗语和目标风险语料构建的风险语料，所述目标风险语料包含与所述第二暗语具有预设关联关系的风险词语；风险确定模块，用于基于所述目标对象和所述目标语料的相似度，以及所述目标语料的风险标签，确定所述目标对象是否为存在风险。

[07]第三方面，本说明书实施例提供了一种数据处理设备，所述数据处理设备包括：处理器；以及被安排成存储计算机可执行指令的存储器，所述可执行指令在被执行时使所述处理器：获取待识别的目标对象；若所述目标对象中包含与第一暗语匹配的词语，则从预先构建的语料库包含的语料中，获取与所述目标对象对应的目标语料，所述预先构建的语料库包括第一语料，所述第一语料为基于第二暗语和目标风险语料构建的风险语料，所述目标风险语料包含与所述第二暗语具有预设关联关系的风险词语；基于所述目标对象和所述目标语料的相似度，以及所述目标语料的风险标签，确定所述目标对象是否为存在风险。

[08]第四方面，本说明书实施例提供了一种存储介质，所述存储介质用于存储计算机可执行指令，所述可执行指令在被执行时实现以下流程：获取待识别的目标对象；若所述目标对象中包含与第一暗语匹配的词语，则从预先构建的语料库包含的语料中，获取与所述目标对象对应的目标语料，所述预先构建的语料库包括第一语料，所述第一语料为基于第二暗语和目标风险语料构建的风险语料，所述目标风险语料包含与所述第二暗语

具有预设关联关系的风险词语；基于所述目标对象和所述目标语料的相似度，以及所述目标语料的风险标签，确定所述目标对象是否为存在风险。

附图说明

[09]为了更清楚地说明本说明书实施例中的技术方案，下面将对实施例描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本说明书中记载的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

[10]图 1A 为本说明书一种数据处理方法实施例的流程图；

[11]图 1B 为本说明书一种数据处理方法的处理过程示意图；

[12]图 2 为本说明书又一种数据处理方法的处理过程示意图；

[13]图 3 为本说明书一种预设风险词语知识图谱的示意图；

[14]图 4 为本说明书一种数据处理方法的处理过程的示意图；

[15]图 5 为本说明书一种数据处理装置实施例的结构示意图；

[16]图 6 为本说明书一种数据处理设备的结构示意图。

具体实施方式

[17]本说明书实施例提供一种数据处理方法、装置及设备。

[18]为了使本技术领域的人员更好地理解本说明书中的技术方案，下面将结合本说明书实施例中的附图，对本说明书实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本说明书一部分实施例，而不是全部的实施例。基于本说明书中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都应当属于本说明书保护的范围。

实施例一

[19]如图 1A 和 1B 所示，本说明书实施例提供一种数据处理方法，该方法的执行主体可以为服务器，其中，该服务器可以是独立的服务器，也可以是由多个服务器组成的服务器集群。该方法具体可以包括以下步骤 S102 至 S106。

[20]在 S102 中，获取待识别的目标对象。

[21]其中，待识别的目标对象可以是任意文本对象、图片对象、视频对象或语音对象等。

[22]在实施中，随着计算机技术的快速发展，网络服务市场的规模日趋庞大，但网络服务的不断发展，也为恶意第三方提供了新的平台，恶意第三方可以通过具有隐蔽含义的暗语绕过风控防控系统进行违法活动，由于具有隐蔽含义的暗语通常与无风险的词语的相似度较高，仅通过词语匹配的方式无法准确的识别出暗语。

[23]可以通过人工基于暗语的上下文，判断当前场景是否存在风险，但待识别的對象的数据量较大，通过人工判断的方式的数据处理效率低、数据处理准确性差，这就导致风险防控的效率和准确率低，因此，需要一种能够在风控场景下，提高针对暗语的风险防控效率和准确率的解决方案。为此，本说明书实施例提供一种可以解决上述问题的技术方案，具体可以参见下述内容。

[24]可以获取风控场景下待识别的目标对象，例如，服务器可以将获取的资源转移业务场景下用户之间的交互内容（如文字交互内容、语音交互内容等）作为待识别的目标对象，具体如，用户 a 可以通过终端设备中安装的资源转移应用程序，触发启动与用户 b 的交互页面，并在该交互页面中与用户 b 进行交互，终端设备可以将用户 a 与用户 b 的交互内容作为待识别的目标对象发送给服务器。

[25]终端设备将交互内容发送给服务器之前，可以通过预设脱敏模型（脱敏模型可以先通过服务器和获取的训练样本进行模型训练后得到）对该交互内容中可能包含的用户隐私数据进行脱敏处理，并将脱敏处理后的交互内容作为目标对象发送给服务器，或者，终端设备在接收到用户的授权指令（即授权终端设备可以将交互内容发送服务器进行风险识别处理）后，可以将交互内容发送给服务器进行处理。

[26]此外，上述是以目标对象为资源转移业务场景下用户之间的交互内容为例，在实际应用场景中，还可以有多种不同的目标对象，例如，目标对象还可以是页面浏览场景下，第三方在预设显示页面投放的对象（如第三方投放的文字内容、图片内容、视频内容等）等，可以根据实际应用场景的不同，确定不同的目标对象，本说明书实施例对目标对象不做具体限定。

[27]在 S104 中，若目标对象中包含与第一暗语匹配的词语，则从预先构建的语料库包含的语料中，获取与目标对象对应的目标语料。

[28]其中，预先构建的语料库可以包括第一语料，第一语料可以为基于第二暗语和目标风险语料构建的风险语料，目标风险语料包含与第二暗语具有预设关联关系的风险词语，第一暗语和第二暗语可以是除词语本身公知的日常的含义外，还包含隐藏含义的词语，第一暗语和第二暗语可以由服务器通过大数据分析等方式确定，例如，“四大件”，该词语本身公知的日常的含义是指四种家用电器，其隐藏含义为恶意第三方窃取用户财产所需的四种隐私数据（如身份证、银行账户、密码和手机号码），第一暗语和第二暗语可以相同，也可以不同，与第二暗语具有预设关联关系的风险词语可以是与第二暗语的相似度大于预设相似度阈值的词语等，目标风险语料可以是包含该风险词语的语料，例如，第二暗语可以为“四大件”，与第二暗语具有预设关联关系的风险词语可以为“三大件”、“必备四样”等，包含该风险词语的目标风险语料可以为“新手入门必备四样”等，可以基于第二暗语“四大件”和目标风险语料“新手入门必备四样”构建第一语料，例如，构建的第一语料可以为“新手入门四大件”等。

[29]在实施中，第一暗语可以有多个，服务器可以根据目标对象的应用场景的场景标识，获取对应的一个或多个第一暗语，再对获取的第一暗语与目标对象进行匹配处理，确定目标对象中是否包含与获取的第一暗语匹配的词语，例如，目标对象对应的应用场景的场景标识为资源转移场景，服务器可以根据该场景标识，获取与资源转移场景对应的第一暗语，在基于正则表达式确定目标对象中是否包含与第一暗语匹配的词语。

[30]此外，由于目标对象的类型可以有多种，如目标对象可以是文本对象、图片对象、视频对象或语音对象中的一个或多个，服务器在进行匹配处理前，若目标对象为非文本对象，则服务器可以对目标对象进行文字提取处理，得到目标对象对应的文本数据，服务器再基于目标对象的文本数据和第一暗语，进行匹配处理，以确定目标对象中是否包含与第一暗语匹配的词语。例如，目标对象可以是语音对象，则服务器可以基于预设语音转换算法，对语音对象进行文本转换处理，得到目标对象的文本数据，或者，目标对象还可以是视频对象，服务器可以对视频对象中的语音数据进行文本转换处理，同时，还可以对视频对象中包含的图片数据进行文字提取处理，并根据处理结果，确定目标对象的文本数据。

[31]若确定目标对象中包含与第一暗语匹配的词语，则服务器可以从预先构建的语料库中，获取与目标对象对应的目标语料。例如，服务器可以基于预设相似度确定模型，依次获取目标对象与语料库中的每个语料的相似度，再将相似度大于第一预设相似度阈值的语料，确定为目标对象对应的目标语料。或者，服务器还可以将语料中包含第一暗语的语料，确定为目标对象对应的目标语料。

[32]上述目标对象的目标语料的确定方法可以有多种，可以根据实际应用场景的不同而有所不同，本说明书实施例对此不作具体限定。

[33]在 S106 中, 基于目标对象和目标语料的相似度, 以及目标语料的风险标签, 确定目标对象是否为存在风险。

[34]在实施中, 服务器可以根据目标对象和每个目标语料的相似度, 确定目标对象与风险标签之前的关联度, 再根据目标对象与风险标签之前的关联度, 确定目标对象是否存在风险。

[35]例如, 假设目标对象对应的目标语料有语料 1、语料 2 和语料 3, 其中, 语料 1 和语料 2 的风险标签为标签 1, 语料 3 的风险标签为标签 2, 假设目标对象和语料 1 的相似度为 70%, 目标对象和语料 2 的相似度为 60%, 目标对象和语料 3 的相似度为 62%, 则服务器可以确定目标对象和标签 1 之间的关联度为 $(70\%+60\%)/2=0.65$, 目标对象和标签 2 之间的关联度为 0.62, 服务器可以将目标对象的风险标签确定为标签 1, 若标签 1 在目标对象对应的应用场景下的风险等级大于预设风险等级, 则可以确定目标对象存在风险。

[36]上述目标对象是否存在风险的确定方法是一种可选地、可实现的确定方法, 在实际应用场景中, 还可以有多种不同的确定方法, 可以根据实际应用场景的不同而有所不同, 本说明书实施例对此不做具体限定。

[37]本说明书实施例提供一种数据处理方法, 通过获取待识别的目标对象, 若目标对象中包含与第一暗语匹配的词语, 则从预先构建的语料库包含的语料中, 获取与目标对象对应的目标语料, 预先构建的语料库包括第一语料, 第一语料为基于第二暗语和目标风险语料构建的风险语料, 目标风险语料包含与第二暗语具有预设关联关系的风险词语, 基于目标对象和目标语料的相似度, 以及目标语料的风险标签, 确定目标对象是否为存在风险, 这样, 可以通过从预先构建的语料库中获取的与目标对象对应的目标语料, 确定目标对象是否存在风险, 避免了人工判断方式导致的数据处理效率低的问题, 另外, 由于预先构建的语料库中的第一语料是基于第二暗语和目标风险语料构建的风险语料, 且目标风险语料包含与第二暗语具有预设关联关系的风险词语, 因而, 通过确定的目标语料与目标对象的相似度, 以及目标语料的风险标签, 可以准确的确定目标对象是否存在风险, 即可以提高风控场景下针对暗语的风险防控效率和准确率。

实施例二

[38]如图 2 所示, 本说明书实施例提供一种数据处理方法, 该方法的执行主体可以为服务器, 其中, 该服务器可以是独立的服务器, 也可以是由多个服务器组成的服务器集群。该方法具体可以包括以下步骤。

[39]在 S102 中, 获取待识别的目标对象。

[40]在 S202 中, 获取包含与第二暗语具有预设关联关系的风险词语的目标风险语料。

[41]在实施中, 在实际应用中, 上述 S202 的处理方式可以多种多样, 以下再提供一种可选的实现方式, 具体可以参见步骤一~步骤三。

[42]步骤一, 获取与第二暗语具有预设关联关系的第一风险词语。

[43]在实施中, 服务器可以通过预先构建的关联度识别模型, 确定第二暗语与风险词名单中每个风险词语之间的关联度, 服务器再基于第二暗语与每个风险词语之间的关联度, 确定与第二暗语具有预设关联关系的第一风险词语。

[44]其中, 关联度识别模型可以是基于历史暗语和历史词语对由深度学习算法构建的模型训练得到。

[45]此外, 第一风险词语的确定方法还可以有多种, 例如, 还可以通过人工的方式确定与第二暗语具有预设关联关系的第一风险词语, 可以根据实际应用场景的不同选取不同的确定方法, 本说明书实施例对此不作具体限定。

[46]步骤二，获取预设风险词语知识图谱中与第一风险词语具有预设关联关系的第二风险词语。

[47]其中，预设风险词语知识图谱可以用于存储风险实体（如用户、业务等）和风险词语等，风险词语知识图谱可以存储在图数据库中，预设风险词语知识图谱可以是服务器基于历史风险词语以及包含历史风险词语的风险语料构建的知识图谱。

[48]在实施中，由于暗语的更新速度较快，模型可能无法准确的识别出暗语包含的隐蔽含义，因此，可以通过人工的方式确定与第二暗语具有预设关联关系的风险词语，而由于风险词语的数量可能较多，为提高数据处理效率，可以通过人工的方式确定与第二暗语具有预设关联关系的第一风险词语，即服务器可以接收由人工确定的与第二暗语具有预设关联关系的第一风险词语，服务器再基于预设风险词语知识图谱，获取与第一风险词语具有预设关联关系的第二风险词语。

[49]例如，假设第二暗语为“信封号”（该词语的本身公知的日常的含义为信封的尺寸，如大号、中号、小号等，其隐蔽含义为用户的账号，如手机号码、即时通讯账号、资源转移账号等），与第二暗语具有预设关联关系的第一风险词语可以为“信件号”，服务器可以在预设风险词语知识图谱中，查询与“信件号”具有预设关联关系的第二风险词语，如可以根据预设风险词语知识图谱中风险词语的关联关系，获取对应的第二风险词语，具体如图3所示的风险词语知识图谱中，根据该风险词语知识图谱，可以获取如“信箱号”、“储物柜编号”、“邮箱容量”等第二风险词语。

[50]上述获取第二风险词语的获取方法是一种可选地、可实现的获取方法，在实际应用场景中，还可以有多种不同的获取方法，可以根据实际应用场景的不同而有所不同，本说明书实施例对此不作具体限定。

[51]步骤三，将包含第一风险词语的风险语料，以及包含第二风险词语的风险语料，确定为目标风险语料。

[52]在实施中，包含第一风险词语（或第二风险词语）的风险语料可以有多个，包含风险词语（即第一风险词语或第二风险词语）的风险语料可以是基于预设风险识别模型获取的风险语料。

[53]在S204中，基于第二暗语，对目标风险语料中的风险词语进行替换处理，得到第一语料。

[54]在实施中，例如，目标风险语料可以为“取走储物柜编号为xx内的所有有价值的物品”，服务器可以将该目标风险语料中的“储物柜编号”（即与第二暗语具有预设关联关系的风险词语）替换为第二暗语“信封号”，替换处理得到的第一语料可以为“取走信封号为xx内的所有有价值的物品”。

[55]由于暗语的更新速度较快，能够获取的包含暗语的风险语料的数据量较小，通过上述方式就可以构建较多的包含暗语的风险语料（即第一语料），且用于构建第一风险语料的目标风险语料是已标注为风险的语料，因此可以通过构建的第一语料，提高后续基于构建的第一语料进行暗语识别的识别准确性。

[56]在S206中，基于第一语料，构建语料库。

[57]其中，预先构建的语料库还可以包括第二语料，第二语料可以为包含第二暗语的无风险语料。

[58]在实施中，服务器可以获取包含第二暗语的无风险语料，并将获取的无风险语料确定为第二语料，例如，第二暗语可以为“信封号”，获取的第二语料可以为“购买若干数量信封号为小号的信封”等。

[59]服务器可以将包含第二暗语的无风险语料确定为第二语料，并基于第一语料和第二

语料, 构建语料库。这样, 构建的语料库即为包含黑样本(即第一语料)和白样本(即第二语料)的语料库, 基于该语料库可以提高针对暗语的识别准确性。

[60]在实际应用中, 上述基于第一语料和第二语料, 构建语料库的处理方式可以多种多样, 以下再提供一种可选的实现方式, 具体可以参见步骤一~步骤二。

[61]步骤一, 基于预先训练的向量提取模型, 对第一语料和第二语料进行特征提取处理, 得到与第一语料对应的第一表征向量, 以及与第二语料对应的第二表征向量。

[62]其中, 向量提取模型可以为能够对语料进行特征提取的模型, 例如, 向量提取模型可以为基于 Transformer 的双向编码器表示(Bidirectional Encoder Representations from Transformer, BERT)模型, 也可以是风险识别场景下的分类模型中的向量提取子模型, 向量提取模型的构成方式可以有多种, 可以根据实际应用场景的不同, 选取不同的向量提取模型, 本说明书实施例对此不做具体限定。

[63]步骤二, 基于第二暗语、第一表征向量和第一语料的风险标签, 以及第二表征向量和第二语料的风险标签, 构建语料库。

[64]在 S208 中, 基于预先训练的向量提取模型, 对目标对象进行特征提取处理, 得到与目标对象对应的目标表征向量。

[65]其中, 向量提取模型可以为上述 S206 中的向量提取模型, 即对目标对象进行特征提取处理的模型可以与对第一语料(或第二语料)进行特征提取处理的模型相同。

[66]在 S210 中, 基于第一暗语和第二暗语之间的相似度, 和/或目标表征向量与语料库中的表征向量之间的相似度, 获取与目标对象对应的目标语料。

[67]在实施中, 目标语料可以有多个, 如图 4 所示, 在离线阶段, 服务器可以基于第二暗语和风险知识图谱构建语料库, 并在在线阶段, 获取待识别的目标对象, 服务器可以基于风险词名单对待识别的目标对象进行匹配处理, 以得到针对目标对象的匹配结果, 若服务器基于匹配结果, 确定目标对象中包含与第一暗语匹配的词语, 则服务器可以基于预先训练的向量提取模型, 对目标对象进行特征提取处理, 得到与目标对象对应的目标表征向量, 然后在获取第一暗语和第二暗语之间的相似度、目标表征向量与语料库中的表征向量之间的相似度, 以得到与目标对象对应的目标语料。

[68]其中, 在确定第一暗语后, 服务器可以将第一暗语添加至风险词名单, 即服务器可以对风险词名单进行实时更新, 这样, 服务器就可以基于预设时间级(如小时级)对目标对象进行风险识别处理。

[69]在 S212 中, 获取目标对象的目标表征向量和目标语料的表征向量之间的相似度, 对目标语料进行排序。

[70]在实施中, 服务器在获取到目标语料后, 可以基于目标对象的目标表征向量和目标语料的表征向量之间的相似度, 对目标语料进行排序, 如可以按照从大到小的顺序对目标对象的目标表征向量和目标语料的表征向量之间的相似度进行排序。

[71]在 S214 中, 基于目标语料的排序顺序, 以及目标语料的风险标签, 确定目标对象是否为存在风险。

[72]在实施中, 可以基于目标语料的排序顺序, 以及目标语料的风险标签, 确定目标对象的风险值。

[73]在实际应用中, 上述目标对象的风险值的确定方式可以多种多样, 以下再提供一种可选的实现方式, 具体可以参见步骤一~步骤三。

[74]步骤一, 基于目标语料的排序顺序, 确定每个目标语料的风险权重。

[75]步骤二, 基于目标语料的风险标签, 确定目标语料的风险值。

[76]在实施中，由于不同的应用场景的风险识别需求不同，例如，在资源转移场景下，风险标签 1 的风险值可以为 0.2，在即时通讯场景下，风险标签的风险值可以为 0.5，因此，服务器可以根据目标语料的应用场景，获取风险标签对应的风险值，在确定目标语料的风险值。

[77]此外，为区分第一语料和第二语料（即区分风险语料和无风险语料），服务器可以将第一语料的风险标签对应的风险值设置为正数，将第二语料的风险标签对应的风险值设置为负数。

[78]步骤三，基于每个目标语料的风险权重和风险值，确定目标对象的风险值。

[79]步骤四，基于目标对象的风险值，确定目标对象是否存在风险。

[80]在实施中，服务器可以基于每个目标语料的排序，确定目标语料的风险权重（如目标语料有 10 个，则第 1 个目标语料的风险权重可以为 $(10-1)/10=0.9$ ），服务器可以将目标语料的风险权重和风险值的乘积，作为该目标语料的目标风险值，最后，服务器可以将目标语料的目标风险值的平均值（或最大值等），确定为目标对象的风险值。

[81]例如，目标语料可以有语料 a、语料 b 和语料 c，目标对象的目标表征向量和目标语料的表征向量之间的相似度、基于目标语料的风险标签确定目标语料的风险值可以如下表 1 所示。

表 1

	相似度	排序	风险权重	风险值	目标风险值
语料 a	50%	3	$1/6=0.17$	3（风险标签为标签 1）	0.51
语料 b	60%	1	$3/6=0.5$	-1（风险标签为无风险的标签 2）	-0.5
语料 c	55%	2	$2/6=0.33$	2（风险标签为标签 3）	0.66

[82]其中，由于语料库中还包括第二语料（即无风险语料），因此，目标语料也可能包含无风险语料，即目标语料的目标风险值可以为负数，因此，服务器可以将目标对象的目标风险值的和值，确定为目标对象的风险值，或者，服务器还可以基于目标语料的目标风险值的绝对值，确定目标对象的风险值（如可以将绝对值最大的目标风险值确定为目标对象的风险值）。

[83]上述目标对象的风险值的确定方法是一种可选地、可实现的确定方法，在实际应用场景中，还可以有多种不同的确定方法，可以根据实际应用场景的不同选取不同的确定方法，本说明书实施例对此不作具体限定。

[84]服务器在确定目标对象的风险值后，可以基于预设风险阈值，确定目标对象是否存在风险，例如，若目标对象的风险值大于预设风险阈值，则可以确定目标对象存在风险。

[85]若确定目标对象存在风险，服务器可以停止触发与目标对象相关的业务，例如，目标对象可以是终端设备获取的资源转移场景下的用户交互内容，服务器若确定该交互内容存在风险，可以停止触发对应的资源转移业务。

[86]此外，若仅将暗语作为关键词对目标对象进行匹配处理确定目标对象是否存在风险，虽然覆盖效果较好，时效较快，但误召回率较高，而本申请可以在保证高时效的前提下，通过对目标语料的处理，提高对目标对象的风险识别的准确性；而在通过模型识别目标对象是否存在风险的方式中，由于暗语更新速度快，能够获取到的样本数据量有限，且模型的迭代要求较高，对风险响应的时效较差，而本申请可以通过离线阶段构建语料库，保证语料库的样本数据量要求和样本准确性要求，在通过对目标语料的排序等处理，可以提高实时识别的效果，无需对模型进行重新训练和迭代。

[87]本说明书实施例提供一种数据处理方法，获取待识别的目标对象，若目标对象中包

含与第一暗语匹配的词语,则从预先构建的语料库包含的语料中,获取与目标对象对应的目标语料,预先构建的语料库包括第一语料,第一语料为基于第二暗语和目标风险语料构建的风险语料,目标风险语料包含与第二暗语具有预设关联关系的风险词语,基于目标对象和目标语料的相似度,以及目标语料的风险标签,确定目标对象是否为存在风险,这样,可以通过从预先构建的语料库中获取的与目标对象对应的目标语料,确定目标对象是否存在风险,避免了人工判断方式导致的数据处理效率低的问题,另外,由于预先构建的语料库中的第一语料是基于第二暗语和目标风险语料构建的风险语料,且目标风险语料包含与第二暗语具有预设关联关系的风险词语,因而,通过确定的目标语料与目标对象的相似度,以及目标语料的风险标签,可以准确的确定目标对象是否存在风险,即可以提高风控场景下针对暗语的风险防控效率和准确率。

实施例三

[88]以上为本说明书实施例提供的数据处理方法,基于同样的思路,本说明书实施例还提供一种数据处理装置,如图5所示。

[89]该数据处理装置包括:对象获取模块501、语料获取模块502和风险确定模块503,其中:对象获取模块501,用于获取待识别的目标对象;语料获取模块502,用于若所述目标对象中包含与第一暗语匹配的词语,则从预先构建的语料库包含的语料中,获取与所述目标对象对应的目标语料,所述预先构建的语料库包括第一语料,所述第一语料为基于第二暗语和目标风险语料构建的风险语料,所述目标风险语料包含与所述第二暗语具有预设关联关系的风险词语;风险确定模块503,用于基于所述目标对象和所述目标语料的相似度,以及所述目标语料的风险标签,确定所述目标对象是否为存在风险。

[90]本说明书实施例中,所述装置还包括:第一获取模块,用于获取包含与所述第二暗语具有预设关联关系的风险词语的目标风险语料;构建模块,用于基于所述第二暗语,对所述目标风险语料中的风险词语进行替换处理,得到所述第一语料,并基于所述第一语料,构建所述语料库。

[91]本说明书实施例中,所述第一获取模块,用于:获取与所述第二暗语具有预设关联关系的第一风险词语;获取预设风险词语知识图谱中与所述第一风险词语具有所述预设关联关系的第二风险词语;将包含所述第一风险词语的风险语料,以及包含所述第二风险词语的风险语料,确定为所述目标风险语料。

[92]本说明书实施例中,所述预先构建的语料库还包括第二语料,所述第二语料为包含所述第二暗语的无风险语料,所述构建模块,用于:将包含所述第二暗语的无风险语料确定为所述第二语料,并基于所述第一语料和所述第二语料,构建所述语料库。

[93]本说明书实施例中,所述构建模块,用于:基于预先训练的向量提取模型,对所述第一语料和所述第二语料进行特征提取处理,得到与所述第一语料对应的第一表征向量,以及与所述第二语料对应的第二表征向量;基于所述第二暗语、所述第一表征向量和所述第一语料的风险标签,以及所述第二表征向量和所述第二语料的风险标签,构建所述语料库;所述语料获取模块502,用于:基于所述预先训练的向量提取模型,对所述目标对象进行特征提取处理,得到与所述目标对象对应的目标表征向量;基于所述第一暗语和所述第二暗语之间的相似度,和/或所述目标表征向量与所述语料库中的表征向量之间的相似度,获取与所述目标对象对应的目标语料。

[94]本说明书实施例中,所述目标语料有多个,所述风险确定模块503,用于:获取所述目标对象的目标表征向量和所述目标语料的表征向量之间的相似度,对所述目标语料进行排序;基于所述目标语料的排序顺序,以及所述目标语料的风险标签,确定所述目标对象是否为存在风险。

[95]本说明书实施例中,所述风险确定模块503,用于:基于所述目标语料的排序顺序,

以及所述目标语料的风险标签，确定所述目标对象的风险值，并基于所述目标对象的风险值，确定所述目标对象是否存在风险。

[96]本说明书实施例中，所述风险确定模块 503，用于：基于所述目标语料的排序顺序，确定每个所述目标语料的风险权重；基于所述目标语料的风险标签，确定所述目标语料的风险值；基于每个所述目标语料的风险权重和风险值，确定所述目标对象的风险值。

[97]本说明书实施例提供一种数据处理装置，获取待识别的目标对象，若目标对象中包含与第一暗语匹配的词语，则从预先构建的语料库包含的语料中，获取与目标对象对应的目标语料，预先构建的语料库包括第一语料，第一语料为基于第二暗语和目标风险语料构建的风险语料，目标风险语料包含与第二暗语具有预设关联关系的风险词语，基于目标对象和目标语料的相似度，以及目标语料的风险标签，确定目标对象是否为存在风险，这样，可以通过从预先构建的语料库中获取的与目标对象对应的目标语料，确定目标对象是否存在风险，避免了人工判断方式导致的数据处理效率低的问题，另外，由于预先构建的语料库中的第一语料是基于第二暗语和目标风险语料构建的风险语料，且目标风险语料包含与第二暗语具有预设关联关系的风险词语，因而，通过确定的目标语料与目标对象的相似度，以及目标语料的风险标签，可以准确的确定目标对象是否存在风险，即可以提高风控场景下针对暗语的风险防控效率和准确率。

实施例四

[98]基于同样的思路，本说明书实施例还提供一种数据处理设备，如图 6 所示。

[99]数据处理设备可因配置或性能不同而产生比较大的差异，可以包括一个或一个以上的处理器 601 和存储器 602，存储器 602 中可以存储有一个或一个以上存储应用程序或数据。其中，存储器 602 可以是短暂存储或持久存储。存储在存储器 602 的应用程序可以包括一个或一个以上模块（图示未示出），每个模块可以包括对数据处理设备中的一系列计算机可执行指令。更进一步地，处理器 601 可以设置为与存储器 602 通信，在数据处理设备上执行存储器 602 中的一系列计算机可执行指令。数据处理设备还可以包括一个或一个以上电源 603，一个或一个以上有线或无线网络接口 604，一个或一个以上输入输出接口 605，一个或一个以上键盘 606。

[100]具体在本实施例中，数据处理设备包括有存储器，以及一个或一个以上的程序，其中一个或者一个以上程序存储于存储器中，且一个或者一个以上程序可以包括一个或一个以上模块，且每个模块可以包括对数据处理设备中的一系列计算机可执行指令，且经配置以由一个或者一个以上处理器执行该一个或者一个以上程序包含用于进行以下计算机可执行指令：获取待识别的目标对象；若所述目标对象中包含与第一暗语匹配的词语，则从预先构建的语料库包含的语料中，获取与目标对象对应的目标语料，所述预先构建的语料库包括第一语料，所述第一语料为基于第二暗语和目标风险语料构建的风险语料，所述目标风险语料包含与第二暗语具有预设关联关系的风险词语；基于所述目标对象和所述目标语料的相似度，以及所述目标语料的风险标签，确定所述目标对象是否为存在风险。

[101]可选地，在所述从预先构建的语料库包含的语料中，获取与目标对象对应的目标语料之前，还包括：获取包含与第二暗语具有预设关联关系的风险词语的目标风险语料；基于第二暗语，对所述目标风险语料中的风险词语进行替换处理，得到所述第一语料，并基于所述第一语料，构建所述语料库。

[102]可选地，所述获取包含与第二暗语具有预设关联关系的风险词语的目标风险语料，包括：获取与第二暗语具有预设关联关系的第一风险词语；获取预设风险词语知识图谱中与第一风险词语具有所述预设关联关系的第二风险词语；将包含所述第一风险词语的风险语料，以及包含所述第二风险词语的风险语料，确定为所述目标风险语料。

[103]可选地,所述预先构建的语料库还包括第二语料,所述第二语料为包含所述第二暗语的无风险语料,所述基于所述第一语料,构建所述语料库,包括:将包含所述第二暗语的无风险语料确定为所述第二语料,并基于所述第一语料和所述第二语料,构建所述语料库。

[104]可选地,所述基于所述第一语料和所述第二语料,构建所述语料库,包括:基于预先训练的向量提取模型,对所述第一语料和所述第二语料进行特征提取处理,得到与所述第一语料对应的第一表征向量,以及与所述第二语料对应的第二表征向量;基于所述第二暗语、所述第一表征向量和所述第一语料的风险标签,以及所述第二表征向量和所述第二语料的风险标签,构建所述语料库;所述从预先构建的语料库包含的语料中,获取与所述目标对象对应的目标语料,包括:基于所述预先训练的向量提取模型,对所述目标对象进行特征提取处理,得到与所述目标对象对应的目标表征向量;基于所述第一暗语和所述第二暗语之间的相似度,和/或所述目标表征向量与所述语料库中的表征向量之间的相似度,获取与所述目标对象对应的目标语料。

[105]可选地,所述目标语料有多个,所述基于所述目标对象和所述目标语料的相似度,以及所述目标语料的风险标签,确定所述目标对象是否为存在风险,包括:获取所述目标对象的目标表征向量和所述目标语料的表征向量之间的相似度,对所述目标语料进行排序;基于所述目标语料的排序顺序,以及所述目标语料的风险标签,确定所述目标对象是否为存在风险。

[106]可选地,所述基于所述目标语料的排序顺序,以及所述目标语料的风险标签,确定所述目标对象是否为存在风险,包括:基于所述目标语料的排序顺序,以及所述目标语料的风险标签,确定所述目标对象的风险值,并基于所述目标对象的风险值,确定所述目标对象是否存在风险。

[107]可选地,所述基于所述目标语料的排序顺序,以及所述目标语料的风险标签,确定所述目标对象的风险值,包括:基于所述目标语料的排序顺序,确定每个所述目标语料的风险权重;基于所述目标语料的风险标签,确定所述目标语料的风险值;基于每个所述目标语料的风险权重和风险值,确定所述目标对象的风险值。

[108]本说明书实施例提供一种数据处理设备,获取待识别的目标对象,若目标对象中包含与第一暗语匹配的词语,则从预先构建的语料库包含的语料中,获取与目标对象对应的目标语料,预先构建的语料库包括第一语料,第一语料为基于第二暗语和目标风险语料构建的风险语料,目标风险语料包含与第二暗语具有预设关联关系的风险词语,基于目标对象和目标语料的相似度,以及目标语料的风险标签,确定目标对象是否为存在风险,这样,可以通过从预先构建的语料库中获取的与目标对象对应的目标语料,确定目标对象是否存在风险,避免了人工判断方式导致的数据处理效率低的问题,另外,由于预先构建的语料库中的第一语料是基于第二暗语和目标风险语料构建的风险语料,且目标风险语料包含与第二暗语具有预设关联关系的风险词语,因而,通过确定的目标语料与目标对象的相似度,以及目标语料的风险标签,可以准确的确定目标对象是否存在风险,即可以提高风控场景下针对暗语的风险防控效率和准确率。

实施例五

[109]本说明书实施例还提供一种计算机可读存储介质,计算机可读存储介质上存储有计算机程序,该计算机程序被处理器执行时实现上述数据处理方法实施例的各个过程,且能达到相同的技术效果,为避免重复,这里不再赘述。其中,所述的计算机可读存储介质,如只读存储器(Read-Only Memory,简称ROM)、随机存取存储器(Random Access Memory,简称RAM)、磁碟或者光盘等。

[110]本说明书实施例提供一种计算机可读存储介质,获取待识别的目标对象,若目标对象中包含与第一暗语匹配的词语,则从预先构建的语料库包含的语料中,获取与目标对

象对应的目标语料,预先构建的语料库包括第一语料,第一语料为基于第二暗语和目标风险语料构建的风险语料,目标风险语料包含与第二暗语具有预设关联关系的风险词语,基于目标对象和目标语料的相似度,以及目标语料的风险标签,确定目标对象是否存在风险,这样,可以通过从预先构建的语料库中获取的与目标对象对应的目标语料,确定目标对象是否存在风险,避免了人工判断方式导致的数据处理效率低的问题,另外,由于预先构建的语料库中的第一语料是基于第二暗语和目标风险语料构建的风险语料,且目标风险语料包含与第二暗语具有预设关联关系的风险词语,因而,通过确定的目标语料与目标对象的相似度,以及目标语料的风险标签,可以准确的确定目标对象是否存在风险,即可以提高风控场景下针对暗语的风险防控效率和准确率。

[111]上述对本说明书特定实施例进行了描述。其它实施例在所附权利要求书的范围内。在一些情况下,在权利要求书中记载的动作或步骤可以按照不同于实施例中的顺序来执行并且仍然可以实现期望的结果。另外,在附图中描绘的过程不一定要求示出的特定顺序或者连续顺序才能实现期望的结果。在某些实施方式中,多任务处理和并行处理也是可以的或者可能是有利的。

[112]在20世纪90年代,对于一个技术的改进可以很明显地区分是硬件上的改进(例如,对二极管、晶体管、开关等电路结构的改进)还是软件上的改进(对于方法流程的改进)。然而,随着技术的发展,当今的很多方法流程的改进已经可以视为硬件电路结构的直接改进。设计人员几乎都通过将改进的方法流程编程到硬件电路中来得到相应的硬件电路结构。因此,不能说一个方法流程的改进就不能用硬件实体模块来实现。例如,可编程逻辑器件(Programmable Logic Device, PLD)(例如现场可编程门阵列(Field Programmable Gate Array, FPGA))就是这样一种集成电路,其逻辑功能由用户对器件编程来确定。由设计人员自行编程来把一个数字系统“集成”在一片PLD上,而不需要请芯片制造厂商来设计和制作专用的集成电路芯片。而且,如今,取代手工地制作集成电路芯片,这种编程也多半改用“逻辑编译器(logic compiler)”软件来实现,它与程序开发撰写时所用的软件编译器相类似,而要编译之前的原始代码也得用特定的编程语言来撰写,此称之为硬件描述语言(Hardware Description Language, HDL),而HDL也并非仅有一种,而是有许多种,如ABEL(Advanced Boolean Expression Language)、AHDL(Altera Hardware Description Language)、Confluence、CUPL(Cornell University Programming Language)、HDCal、JHDL(Java Hardware Description Language)、Lava、Lola、MyHDL、PALASM、RHDL(Ruby Hardware Description Language)等,目前最普遍使用的是VHDL(Very-High-Speed Integrated Circuit Hardware Description Language)与Verilog。本领域技术人员也应该清楚,只需要将方法流程用上述几种硬件描述语言稍作逻辑编程并编程到集成电路中,就可以很容易得到实现该逻辑方法流程的硬件电路。

[113]控制器可以按任何适当的方式实现,例如,控制器可以采取例如微处理器或处理器以及存储可由该(微)处理器执行的计算机可读程序代码(例如软件或固件)的计算机可读介质、逻辑门、开关、专用集成电路(Application Specific Integrated Circuit, ASIC)、可编程逻辑控制器和嵌入微控制器的形式,控制器的例子包括但不限于以下微控制器:ARC 625D、Atmel AT91SAM、Microchip PIC18F26K20 以及Silicone Labs C8051F320,存储器控制器还可以被实现为存储器的控制逻辑的一部分。本领域技术人员也知道,除了以纯计算机可读程序代码方式实现控制器以外,完全可以通过将方法步骤进行逻辑编程来使得控制器以逻辑门、开关、专用集成电路、可编程逻辑控制器和嵌入微控制器等的形式来实现相同功能。因此这种控制器可以被认为是一种硬件部件,而对其内包括的用于实现各种功能的装置也可以视为硬件部件内的结构。或者甚至,可以将用于实现各种功能的装置视为既可以是实现方法的软件模块又可以是硬件部件内的结构。

[114]上述实施例阐明的系统、装置、模块或单元,具体可以由计算机芯片或实体实现,或者由具有某种功能的产品来实现。一种典型的实现设备为计算机。具体的,计算机例

如可以为个人计算机、膝上型计算机、蜂窝电话、相机电话、智能电话、个人数字助理、媒体播放器、导航设备、电子邮件设备、游戏控制台、平板计算机、可穿戴设备或者这些设备中的任何设备的组合。

[115]为了描述的方便，描述以上装置时以功能分为各种单元分别描述。当然，在实施本说明书一个或多个实施例时可以把各单元的功能在同一个或多个软件和/或硬件中实现。

[116]本领域内的技术人员应明白，本说明书的实施例可提供为方法、系统、或计算机程序产品。因此，本说明书一个或多个实施例可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且，本说明书一个或多个实施例可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质（包括但不限于磁盘存储器、CD-ROM、光学存储器等）上实施的计算机程序产品的形式。

[117]本说明书的实施例是参照根据本说明书实施例的方法、设备（系统）、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器，使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[118]这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中，使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品，该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[119]这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上，使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理，从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[120]在一个典型的配置中，计算设备包括一个或多个处理器（CPU）、输入/输出接口、网络接口和内存。

[121]内存可能包括计算机可读介质中的非永久性存储器，随机存取存储器（RAM）和/或非易失性内存等形式，如只读存储器（ROM）或闪存（flash RAM）。内存是计算机可读介质的示例。

[122]计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括，但不限于相变内存（PRAM）、静态随机存取存储器（SRAM）、动态随机存取存储器（DRAM）、其他类型的随机存取存储器（RAM）、只读存储器（ROM）、电可擦除可编程只读存储器（EEPROM）、快闪记忆体或其他内存技术、只读光盘只读存储器（CD-ROM）、数字多功能光盘（DVD）或其他光学存储、磁盒式磁带，磁带磁盘存储或其他磁性存储设备或任何其他非传输介质，可用于存储可以被计算设备访问的信息。按照本文中的界定，计算机可读介质不包括暂存电脑可读媒体（transitory media），如调制的数据信号和载波。

[123]还需要说明的是，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、商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、商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、商品或者设备中还存在另外的相同要素。

[124]本领域技术人员应明白,本说明书的实施例可提供为方法、系统或计算机程序产品。因此,本说明书一个或多个实施例可采用完全硬件实施例、完全软件实施例或结合软件和硬件方面的实施例的形式。而且,本说明书一个或多个实施例可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[125]本说明书一个或多个实施例可以在由计算机执行的计算机可执行指令的一般上下文中描述,例如程序模块。一般地,程序模块包括执行特定任务或实现特定抽象数据类型的例程、程序、对象、组件、数据结构等等。也可以在分布式计算环境中实践本说明书一个或多个实施例,在这些分布式计算环境中,由通过通信网络而被连接的远程处理设备来执行任务。在分布式计算环境中,程序模块可以位于包括存储设备在内的本地和远程计算机存储介质中。

[126]本说明书中的各个实施例均采用递进的方式描述,各个实施例之间相同相似的部分互相参见即可,每个实施例重点说明的都是与其他实施例的不同之处。尤其,对于系统实施例而言,由于其基本相似于方法实施例,所以描述的比较简单,相关之处参见方法实施例的部分说明即可。

[127]以上所述仅为本说明书的实施例而已,并不用于限制本说明书。对于本领域技术人员来说,本说明书可以有各种更改和变化。凡在本说明书的精神和原理之内所作的任何修改、等同替换、改进等,均应包含在本说明书的权利要求范围之内。

权利要求书

1、一种数据处理方法，包括：

获取待识别的目标对象；

若所述目标对象中包含与第一暗语匹配的词语，则从预先构建的语料库包含的语料中，获取与所述目标对象对应的目标语料，所述预先构建的语料库包括第一语料，所述第一语料为基于第二暗语和目标风险语料构建的风险语料，所述目标风险语料包含与所述第二暗语具有预设关联关系的风险词语；

基于所述目标对象和所述目标语料的相似度，以及所述目标语料的风险标签，确定所述目标对象是否为存在风险。

2、根据权利要求1所述的方法，在所述从预先构建的语料库包含的语料中，获取与所述目标对象对应的目标语料之前，还包括：

获取包含与所述第二暗语具有预设关联关系的风险词语的目标风险语料；

基于所述第二暗语，对所述目标风险语料中的风险词语进行替换处理，得到所述第一语料，并基于所述第一语料，构建所述语料库。

3、根据权利要求2所述的方法，所述获取包含与所述第二暗语具有预设关联关系的风险词语的目标风险语料，包括：

获取与所述第二暗语具有预设关联关系的第一风险词语；

获取预设风险词语知识图谱中与所述第一风险词语具有所述预设关联关系的第二风险词语；

将包含所述第一风险词语的风险语料，以及包含所述第二风险词语的风险语料，确定为所述目标风险语料。

4、根据权利要求3所述的方法，所述预先构建的语料库还包括第二语料，所述第二语料为包含所述第二暗语的无风险语料，所述基于所述第一语料，构建所述语料库，包括：

将包含所述第二暗语的无风险语料确定为所述第二语料，并基于所述第一语料和所述第二语料，构建所述语料库。

5、根据权利要求4所述的方法，所述基于所述第一语料和所述第二语料，构建所述语料库，包括：

基于预先训练的向量提取模型，对所述第一语料和所述第二语料进行特征提取处理，得到与所述第一语料对应的第一表征向量，以及与所述第二语料对应的第二表征向量；

基于所述第二暗语、所述第一表征向量和所述第一语料的风险标签，以及所述第二表征向量和所述第二语料的风险标签，构建所述语料库；

所述从预先构建的语料库包含的语料中，获取与所述目标对象对应的目标语料，包括：

基于所述预先训练的向量提取模型，对所述目标对象进行特征提取处理，得到与所述目标对象对应的目标表征向量；

基于所述第一暗语和所述第二暗语之间的相似度，和/或所述目标表征向量与所述语料库中的表征向量之间的相似度，获取与所述目标对象对应的目标语料。

6、根据权利要求5所述的方法，所述目标语料有多个，所述基于所述目标对象和所述目标语料的相似度，以及所述目标语料的风险标签，确定所述目标对象是否为存在风险，包括：

获取所述目标对象的目标表征向量和所述目标语料的表征向量之间的相似度，对所述目标语料进行排序；

基于所述目标语料的排序顺序，以及所述目标语料的风险标签，确定所述目标对象是否为存在风险。

7、根据权利要求6所述的方法，所述基于所述目标语料的排序顺序，以及所述目标语料的风险标签，确定所述目标对象是否为存在风险，包括：

基于所述目标语料的排序顺序，以及所述目标语料的风险标签，确定所述目标对象

的风险值，并基于所述目标对象的风险值，确定所述目标对象是否存在风险。

8、根据权利要求7所述的方法，所述基于所述目标语料的排序顺序，以及所述目标语料的风险标签，确定所述目标对象的风险值，包括：

基于所述目标语料的排序顺序，确定每个所述目标语料的风险权重；

基于所述目标语料的风险标签，确定所述目标语料的风险值；

基于每个所述目标语料的风险权重和风险值，确定所述目标对象的风险值。

9、一种数据处理装置，包括：

对象获取模块，用于获取待识别的目标对象；

语料获取模块，用于若所述目标对象中包含与第一暗语匹配的词语，则从预先构建的语料库包含的语料中，获取与所述目标对象对应的目标语料，所述预先构建的语料库包括第一语料，所述第一语料为基于第二暗语和目标风险语料构建的风险语料，所述目标风险语料包含与所述第二暗语具有预设关联关系的风险词语；

风险确定模块，用于基于所述目标对象和所述目标语料的相似度，以及所述目标语料的风险标签，确定所述目标对象是否为存在风险。

10、一种数据处理设备，所述数据处理设备包括：

处理器；以及

被安排成存储计算机可执行指令的存储器，所述可执行指令在被执行时使所述处理器：

获取待识别的目标对象；

若所述目标对象中包含与第一暗语匹配的词语，则从预先构建的语料库包含的语料中，获取与所述目标对象对应的目标语料，所述预先构建的语料库包括第一语料，所述第一语料为基于第二暗语和目标风险语料构建的风险语料，所述目标风险语料包含与所述第二暗语具有预设关联关系的风险词语；

基于所述目标对象和所述目标语料的相似度，以及所述目标语料的风险标签，确定所述目标对象是否为存在风险。

11、一种存储介质，所述存储介质用于存储计算机可执行指令，所述可执行指令在被执行时实现以下流程：

获取待识别的目标对象；

若所述目标对象中包含与第一暗语匹配的词语，则从预先构建的语料库包含的语料中，获取与所述目标对象对应的目标语料，所述预先构建的语料库包括第一语料，所述第一语料为基于第二暗语和目标风险语料构建的风险语料，所述目标风险语料包含与所述第二暗语具有预设关联关系的风险词语；

基于所述目标对象和所述目标语料的相似度，以及所述目标语料的风险标签，确定所述目标对象是否为存在风险。