一种数据匹配方法、装置、系统、设备及介质

相关申请的交叉引用

本申请要求在2022年08月04日提交中国专利局、申请号为202210933633.2、申请名称为“一种数据匹配方法、装置、系统、设备及介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及数据处理技术领域，尤其涉及一种数据匹配方法、装置、系统、设备及介质。

背景技术

当前隐私计算技术主要应用于安全求交和联邦学习。所谓安全求交是指识别双方数据的交集，比如，即识别机构 A 和机构 B 的共有用户，其中安全求交也为纵向联邦学习的第一个步骤，也就是说，先对手机号、身份证号、营业执照号等关键信息进行安全求交，然后再进行下一步联合建模等步骤。

相关技术中，为了识别双方数据的交集或者实现双方数据的匹配，常见的安全求交算法包括基于 RSA 加密算法的安全求交算法等。但是目前的安全求交算法只有当双方数据完全相同时，也就是说，当双方数据的数据类型以及数据包含的字符的数量完全相同时，才能匹配成功。但是实际业务中，往往存在很多数据不完全相同时待匹配的使用场景，因此现有技术中的安全求交算法极大限制了其使用场景，影响匹配的业务范围。

发明内容

本申请提供了一种数据匹配方法、装置、系统、设备及介质，用以解决现有技术中的安全求交算法只能在双方数据完全相同时进行安全求交，使用场景有限，影响数据匹配的业务范围的问题。

第一方面，本申请提供了一种数据匹配方法，应用于第一设备，所述方法包括:

将待匹配的第一数据输入到预先训练完成的向量转化模型中，获得所述第一数据对应的第一向量，并获得所述第一向量对应的二进制向量;

通过不经意传输 OT 协议，获取第一中间向量；基于所述第一中间向量以及第一设定算法，获得第一距离中间结果；其中，所述第一中间向量是基于所述第一向量对应的二进制向量、第二向量及第一随机向量生成的，所述第二向量为第二设备将第二数据输入到所述第二设备中的预先训练完成的向量转化模型中获得的，所述第一随机向量为所述第二设备基于所述 OT 协议生成的;

接收所述第二设备发送的基于所述第二向量、所述第一随机向量以及第二设定算法生成的第二距离中间结果；基于所述第一向量、所述第一距离中间结果和所述第二距离中间结果，确定所述第一向量和第二向量的目标距离;

基于所述目标距离以及预设的第一距离阈值，确定所述第一数据以及所述第二数据是否匹配。

第二方面，本申请提供了一种数据匹配方法，应用于第二设备，所述方法包括:

将待匹配的第二数据输入到预先训练完成的向量转化模型中，获得所述第二数据对应的第二向量；并基于不经意传输 OT 协议，生成第一随机向量;

基于所述第二向量、所述第一随机向量以及第二设定算法生成的第二距离中间结果，并将所述第二距离中间结果发送给第一设备，使第一设备基于第一向量、第一距离中间结果和所述第二距离中间结果，确定所述第一向量和第二向量的目标距离；并使所述第一设备基于所述目标距离以及预设的第一距离阈值，确定所述第一数据以及所述第二数据是否匹配；其中，所述第一向量为将第一数据输入到所述第一设备中的预先训练完成的向量转化模型中获得的；所述第一距离中间结果为基于第一中间向量以及第一设定算法获得的，所述第一中间向量为基于所述第一向量对应的二进制向量、所述第二

向量及所述第一随机向量生成的。

第三方面，本申请提供了一种数据匹配装置，应用于第一设备，所述装置包括:

第一获取模块，用于将待匹配的第一数据输入到预先训练完成的向量转化模型中，获得所述第一数据对应的第一向量，并获得所述第一向量对应的二进制向量;

第一处理模块，用于通过不经意传输 OT 协议，获取第一中间向量；基于所述第一中间向量以及第一设定算法，获得第一距离中间结果；其中，所述第一中间向量是基于所述第一向量对应的二进制向量、第二向量及第一随机向量生成的，所述第二向量为第二设备将第二数据输入到所述第二设备中的预先训练完成的向量转化模型中获得的，所述第一随机向量为所述第二设备基于所述 OT 协议生成的;

所述第一处理模块，还用于接收所述第二设备发送的基于所述第二向量、所述第一随机向量以及第二设定算法生成的第二距离中间结果；基于所述第一向量、所述第一距离中间结果和所述第二距离中间结果，确定所述第一向量和第二向量的目标距离;

第一确定模块，用于基于所述目标距离以及预设的第一距离阈值，确定所述第一数据以及所述第二数据是否匹配。

第四方面，本申请提供了一种数据匹配装置，应用于第二设备，所述装置包括:

第二获取模块，用于将待匹配的第二数据输入到预先训练完成的向量转化模型中，获得所述第二数据对应的第二向量；并基于不经意传输 OT 协议，生成第一随机向量;

第二处理模块，用于基于所述第二向量、所述第一随机向量以及第二设定算法生成的第二距离中间结果，并将所述第二距离中间结果发送给第一设备，使第一设备基于第一向量、第一距离中间结果和所述第二距离中间结果，确定所述第一向量和第二向量的目标距离；并使所述第一设备基于所述目标

距离以及预设的第一距离阈值，确定所述第一数据以及所述第二数据是否匹配；其中，所述第一向量为将第一数据输入到所述第一设备中的预先训练完成的向量转化模型中获得的；所述第一距离中间结果为基于第一中间向量以及第一设定算法获得的，所述第一中间向量为基于所述第一向量对应的二进制向量、所述第二向量及所述第一随机向量生成的。

第五方面，本申请提供了一种数据匹配系统，所述系统包括：

第一设备，用于将待匹配的第一数据输入到预先训练完成的向量转化模型中，获得所述第一数据对应的第一向量，并获得所述第一向量对应的二进制向量；

第二设备，用于将待匹配的第二数据输入到预先训练完成的向量转化模型中，获得所述第二数据对应的第二向量；将待匹配的第二数据输入到预先训练完成的向量转化模型中，获得所述第二数据对应的第二向量；并基于不经意传输 OT 协议，生成第一随机向量；

所述第一设备，还用于通过不经意传输 OT 协议，获取第一中间向量；基于所述第一中间向量以及第一设定算法，获得第一距离中间结果；其中，所述第一中间向量是基于所述第一向量对应的二进制向量、第二向量及第一随机向量生成的；

所述第二设备，还用于基于所述第二向量、所述第一随机向量以及第二设定算法生成的第二距离中间结果，并将所述第二距离中间结果发送给所述第一设备；

所述第一设备，还用于基于所述第一向量、所述第一距离中间结果和所述第二距离中间结果，确定所述第一向量和第二向量的目标距离；基于所述目标距离以及预设的第一距离阈值，确定所述第一数据以及所述第二数据是否匹配。

第六方面，本申请提供了一种电子设备，所述电子设备包括处理器和存储器，所述存储器用于存储程序指令，所述处理器用于执行存储器中存储的计算机程序时实现上述任一所述数据匹配方法的步骤。

第七方面，本申请提供了一种计算机可读存储介质，其存储有计算机程序，所述计算机程序被处理器执行时实现上述任一所述数据匹配方法的步骤。

由于在本申请实施例中，可以分别将待匹配的第一数据和第二数据输入到预先训练完成的向量转化模型中，获得该第一数据对应的第一向量以及第二数据对应的第二向量，并获得第一向量对应的二进制向量，可以通过不经意传输 OT 协议，获取基于第一向量对应的二进制向量、第二向量及第一随机向量生成的第一中间向量，并可以基于第一中间向量以及第一设定算法，获得第一距离中间结果，还可以基于第二向量、第一随机向量以及第二设定算法，获得第二距离中间结果，并可以基于第一向量、第一距离中间结果和第二距离中间结果，确定第一向量和第二向量的目标距离，基于该目标距离以及预设的第一距离阈值，确定所述第一数据以及所述第二数据是否匹配，即可以在第一数据和第二数据不完全相同时，也能实现第一数据和第二数据的模糊匹配，拓宽了使用场景，且在进行模糊匹配过程中引入了 OT 协议，实现了安全求交，保证了匹配过程的安全性，且整个匹配的过程中，第一数据以及第二数据均未以原始数据的形式离开过对应的第一设备以及第二设备，保证了第一数据和第二数据的安全性，实现了原始数据不出库也能实现模糊匹配，进一步保证了匹配过程的安全性。

附图说明

为了更清楚地说明本发明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要介绍，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，对于本领域的普通技术人员来讲，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。

图 1 为本申请实施例提供的第一种数据匹配过程示意图；

图 2 为本申请实施例提供的第二种数据匹配过程示意图；

图 3 为本申请实施例提供的第三种数据匹配过程示意图；

图 4 为本申请一些实施例提供的一种数据匹配装置结构示意图；

图 5 为本申请一些实施例提供的另一种数据匹配装置结构示意图；

图 6 为本申请一些实施例提供的一种数据匹配系统结构示意图；

图 7 为本申请一些实施例提供的一种电子设备结构示意图。

具体实施方式

为了保证双方数据不完全相同时也可以进行匹配，拓宽数据匹配的业务范围，本申请实施例提供了一种数据匹配方法、装置、系统、设备及介质。

为使本申请的目的和实施方式更加清楚，下面将结合本申请示例性实施例中的附图，对本申请示例性实施方式进行清楚、完整地描述，显然，描述的示例性实施例仅是本申请一部分实施例，而不是全部的实施例。

需要说明的是，本申请中对于术语的简要说明，仅是为了方便理解接下来描述的实施方式，而不是意图限定本申请的实施方式。除非另有说明，这些术语应当按照其普通和通常的含义理解。

本申请中说明书和权利要求书及上述附图中的术语“第一”、“第二”、“第三”等是用于区别类似或同类的对象或实体，而不必然意味着限定特定的顺序或先后次序，除非另外注明。应该理解这样使用的用语在适当情况下可以互换。

术语“包括”和“具有”以及他们的任何变形，意图在于覆盖但不排他的包含，例如，包含了一系列组件的产品或设备不必限于清楚地列出的所有组件，而是可包括没有清楚地列出的或对于这些产品或设备固有的其它组件。

术语“模块”是指任何已知或后来开发的硬件、软件、固件、人工智能、模糊逻辑或硬件或/和软件代码的组合，能够执行与该元件相关的功能。

最后应说明的是：以上各实施例仅用以说明本申请的技术方案，而非对其限制；尽管参照前述各实施例对本申请进行了详细的说明，本领域的普通技术人员应当理解：其依然可以对前述各实施例所记载的技术方案进行修改，或者对其中部分或者全部技术特征进行等同替换；而这些修改或者替换，并不使相应技术方案的本质脱离本申请各实施例技术方案的范围。

实施例 1:

图 1 为本申请实施例提供的第一种数据匹配过程示意图，该过程包括以下步骤:

S101: 将待匹配的第一数据输入到预先训练完成的向量转化模型中，获得所述第一数据对应的第一向量，并获得所述第一向量对应的二进制向量。

本申请实施例提供的数据匹配方法应用于电子设备（为方便描述，称为第一设备），第一设备可以是智能终端、PC 或者服务器等设备。

为了保证双方数据不完全相同时也能实现模糊匹配，在本申请实施例中，第一设备中部署了预先训练完成的向量转化模型，该预先训练完成的向量转化模型用于获得待匹配的数据对应的向量，且针对不同的数据，该预先训练完成的向量转化模型输出的向量的维度可以相同。

在一种可能的实施方式中，为了获得待匹配的第一数据（为方便描述，将第一设备中保存的待匹配的数据称为第一数据）对应的第一向量（为方便描述，将第一数据对应的向量称为第一向量），可以将该第一数据输入到预先训练完成的向量转化模型中，该预先训练完成的向量转化模型可以输出该第一数据对应的第一向量，可选的，该第一向量中的每个分量可以为数字，即可以通过该预先训练完成的向量转化模型将第一数据进行了量化。示例性的，以第一数据为“某海市某东新区晴某小卖部”为例，可以将该第一数据“某海市某东新区晴某小卖部”输入到预先训练完成的词向量模型中，输出的该第一数据“某海市某东新区晴某小卖部”对应的第一向量可以为（1.0，2.0，1.5，2.0，3.5）。

在一种可能的实施方式中，为了保证双方数据不完全相同时也能实现模糊匹配，获得第一数据对应的第一向量之后，可以继而将第一向量转化为二进制的二进制向量。示例性的，假设第一数据 U1 对应的第一向量为（x_1，x_2，x_3……，x_m），第一向量中包含 m 个第一分量(x_i)，其中，i 为不大于 m 的任一正整数，可以将每个十进制的第一分量(x_i)分别转化为二进制的二进制向量

（二进制的 0-1 序列），即将$(x_i)_{10}$转化为（$a_0a_1a_2 \dots a_j \dots a_p$）$_2$，$(x_i)_{10} =$（$a_0a_1a_2 \dots a_j \dots a_p$）$_2$，其中$(x_i)_{10}$表示十进制的第一分量，（$a_0a_1a_2 \dots a_j \dots a_p$）$_2$表示该第一分量对应的二进制的二进制向量，每个二进制向量中包含p个二进制分量a_j，其中j为不大于p的任一正整数，$a_j \in$（0,1），则$x_i = \sum_{j=0}^{p} 2^{p-j} a_j$。将第一向量中的每个第一分量均转化为相应的二进制向量后，即得到的第一向量对应的二进制向量，第一向量对应的二进制向量可以认为是由每个第一分量对应的二进制向量组成的二进制向量集。

S102: 通过不经意传输 OT 协议，获取第一中间向量；基于所述第一中间向量以及第一设定算法，获得第一距离中间结果；其中，所述第一中间向量是基于所述第一向量对应的二进制向量、第二向量及第一随机向量生成的，所述第二向量为第二设备将第二数据输入到所述第二设备中的预先训练完成的向量转化模型中获得的，所述第一随机向量为所述第二设备基于所述 OT 协议生成的。

在一种可能的实施方式中，为了保证双方数据不完全相同时也能实现模糊匹配以及保证匹配过程的安全性，第一设备可以通过不经意传输（Oblivious Transfer，OT）协议，获取第一中间向量。第一中间向量中包含的每个第一中间分量可以为：$a_jy_i + C_j$。其中，OT 协议可以为 1-out-2 OT 协议，第一设备可以作为 OT 协议的输入方，第二设备可以作为 OT 协议的输出方。C_j为第二设备基于 OT 协议生成的随机向量（随机数）中包含的第一随机分量，为方便描述，将第二设备生成的随机向量称为第一随机向量。在一种可能的实施方式中，为了实现第一数据和第二数据的模糊匹配，第二设备中也部署了预先训练完成的向量转化模型，用于获得待匹配的第二数据（为方便描述，将第二设备中待匹配的数据称为第二数据）对应的第二向量（为方便描述，将第二数据对应的向量称为第二向量）。可选的，可以将第二数据输入到预先训练完成的向量转化模型中，该预先训练完成的向量转化模型可以输出该第二数据对应的第二向量。假设第二数据 U5 对应的第二向量为（$y_1, y, y_3 \dots\dots, y_m$），

第二向量中包含 m 个第二分量y_i，其中，i 为不大于 m 的任一正整数。第一向量中包含的分量的数量与该第二分量中包含的分量的数量相同，即第一向量的长度与第二向量的长度相等，第一向量和第二向量的长度可以均为预设长度 m，m 可以为任意正整数。由于第一数据对应的第一向量的长度与第二数据对应的第二向量的长度相等，因此，即使第一数据与第二数据不相同，也能实现模糊匹配，拓宽了使用场景。

在一种可能的实施方式中，第一中间向量中包含的每个第一中间分量可以为：$a_j y_i + C_j$。可以理解的，当$a_j = 0$时，则第一设备获取的第一中间分量$a_j y_i + C_j$为C_j；当$a_j = 1$时，则第一设备获取到的第一中间分量$a_j y_i + C_j$为$y_i + C_j$。

获得第一中间向量之后，第一设备可以基于第一中间向量以及第一设定算法，获得第一距离中间结果。在一种可能的实施方式中，第一距离中间结果可以是采用以下过程获得的：

针对第一向量中的每个第一分量x_i，可以基于该第一分量对应的每个二进制分量a_j及第二向量中相应的第二分量y_i，确定该第一分量与相应的第二分量的乘积$(x_i\ y_i)$；并可以基于第一随机向量中相应的第一随机分量C_j、及该第一随机分量在相应的二进制向量中的位置信息（如 p-j 等），确定该第一分量对应的第一随机分量子值（如$-2\sum_{j=0}^{p} 2^{p-j} C_j$）。可以基于该第一分量与相应的第二分量的乘积$(x_i\ y_i)$与该第一随机分量子值的和，确定第一距离中间结果。

为方便理解，下面用公式形式对基于第一中间向量以及第一设定算法，确定第一距离中间结果的过程进行解释说明。假设第一距离中间结果用 E(a) 表示，则 $E(a)=\sum_{i=1}^{m}\{-2 * [\sum_{j=0}^{p} 2^{p-j}(a_j y_i + C_j)]\}$

$=\sum_{i=1}^{m}\{-2 * [\sum_{j=0}^{p} 2^{p-j} a_j y_i + \sum_{j=0}^{p} 2^{p-j} C_j]\}$

$=\sum_{i=1}^{m}\{-2 * [x_i y_i + \sum_{j=0}^{p} 2^{p-j} C_j]\}$

$=\sum_{i=1}^{m}(-2x_i y_i - 2\sum_{j=0}^{p} 2^{p-j} C_j)$。

S103：接收所述第二设备发送的基于所述第二向量、所述第一随机向量以及第二设定算法生成的第二距离中间结果；基于所述第一向量、所述第一

距离中间结果和所述第二距离中间结果，确定所述第一向量和第二向量的目标距离。

在一种可能的实施方式中，为了保证双方数据不完全相同时也能实现模糊匹配以及保证匹配过程的安全性，第二设备可以基于第二向量、第一随机向量以及第二设定算法生成的第二距离中间结果。在一种可能的实施方式中，第二距离中间结果可以是采用以下过程获得的:

针对第二向量中的每个第二分量y_i，确定该第二分量的第二平方分量(y_i^2)；并可以基于第一随机向量中相应的第一随机分量C_j、及该第一随机分量在相应的二进制向量中的位置信息（如 p-j 等），确定该第二分量对应的第二随机分量子值（如$\sum_{j=0}^{p} 2^{p-j+1} C_j$）；并可以基于该第二平方分量与该第二随机分量子值的和，确定第二距离中间结果。

为方便理解，下面用公式形式对基于第二向量、第一随机向量以及第二设定算法生成第二距离中间结果的过程进行解释说明。假设第二距离中间结果用 E(b)表示，则：$E(b) = \sum_{i=1}^{m}\left(y_i^2 + \sum_{j=0}^{p} 2^{p-j+1} C_j\right)$。

第二设备可以将第二距离中间结果E(b)发送给第一设备。第一设备可以进而基于第一向量、第一距离中间结果和第二距离中间结果，确定第一向量和第二向量的目标距离（为方便描述，将第一向量和第二向量之间的距离称为目标距离）。可选的，基于第一向量、第一距离中间结果和第二距离中间结果，确定第一向量和第二向量的目标距离的过程可以如下:

第一设备针对第一向量中的每个第一分量，可以先基于该第一分量的第一平方分量x_i^2与第一距离中间结果 E(a)和第二距离中间结果 E(b)的和（$x_i^2 + E(a) + E(b)$），确定该第一分量对应的子距离$(x_i - y_i)^2$，然后基于每个第一分量对应的子距离的和，确定第一向量和第二向量的目标距离。

例如，计算每个第一分量对应的子距离的和的过程可以如下:

$$\sum_{i=1}^{m}\left(x_i^2\right) + E(a) + E(b)$$

$$= \sum_{i=1}^{m}\left(x_i^2\right) + \sum_{i=1}^{m}\left(-2x_i y_i - 2\sum_{j=0}^{p} 2^{p-j} C_j\right) + \sum_{i=1}^{m}\left(y_i^2 + \sum_{j=0}^{p} 2^{p-j+1} C_j\right)$$

$=\sum_{i=1}^{m}(x_i^2)-2\sum_{i=1}^{m}(x_i y_i)+\sum_{i=1}^{m}(y_i^2)$

$=\sum_{i=1}^{m}(x_i-y_i)^2$。

可以基于每个第一分量对应的子距离（$\sum_{i=1}^{m}(x_i-y_i)^2$）以及欧氏距离公式，计算第一向量和第二向量之间的目标距离，示例性的，第一向量和第二向量之间的目标距离$d=\sqrt{\sum_{i=1}^{m}(x_i-y_i)^2}$。

在另外一种可能的实施方式中，还可以基于余弦距离公式或者汉明距离公式等确定第一向量与第二向量之间的距离，在此不再赘述。

S104：基于所述目标距离以及预设的第一距离阈值，确定所述第一数据以及所述第二数据是否匹配。

在一种可能的实施方式中，为了确定第一数据与第二数据是否匹配，确定了第一向量与第二向量之间的目标距离后，可以将该目标距离与预设的第一距离阈值进行比较，根据比较结果，确定该第一数据与该第二数据是否匹配。示例性的，可以判断第一向量与第二向量之间的目标距离是否小于预设的第一距离阈值，如果第一向量与第二向量之间的目标距离小于预设的第一距离阈值，则可以认为第一数据与第二数据匹配。而如果第一向量与第二向量之间的目标距离不小于预设的第一距离阈值，则可以确定第一数据与第二数据不匹配。可选的，预设的第一距离阈值可以为 1，还可以为 1.5 等，本申请对第一距离阈值不作具体限定，可以根据需求灵活设置。其中，第一向量与第二向量之间的目标距离越小，第一向量与第二向量越匹配。

为了确定第二数据与第一数据是否完全匹配，在一种可能的实施方式中，还可以判断第一向量与第二向量之间的目标距离是否等于预设的第二距离阈值，若该目标距离等于预设的第二距离阈值，则可以认为第一数据与第二数据相同，也就是说，该第一数据与该第二数据完全匹配。可选的，该第二距离阈值可以小于第一距离阈值，且第二距离阈值可以为 0。

由于在本申请实施例中，可以分别将待匹配的第一数据和第二数据输入到预先训练完成的向量转化模型中，获得该第一数据对应的第一向量以及第二数据对应的第二向量，并获得第一向量对应的二进制向量，可以通过不经

意传输 OT 协议，获取基于第一向量对应的二进制向量、第二向量及第一随机向量生成的第一中间向量，并可以基于第一中间向量以及第一设定算法，获得第一距离中间结果，还可以基于第二向量、第一随机向量以及第二设定算法，获得第二距离中间结果，并可以基于第一向量、第一距离中间结果和第二距离中间结果，确定第一向量和第二向量的目标距离，基于该目标距离以及预设的第一距离阈值，确定所述第一数据以及所述第二数据是否匹配，即可以在第一数据和第二数据不完全相同时，也能实现第一数据和第二数据的模糊匹配，拓宽了使用场景，且在进行模糊匹配过程中引入了 OT 协议，实现了安全求交，保证了匹配过程的安全性，且整个匹配的过程中，第一数据以及第二数据均未以原始数据的形式离开过对应的第一设备以及第二设备，保证了第一数据和第二数据的安全性，实现了原始数据不出库也能实现模糊匹配，进一步保证了匹配过程的安全性。

实施例 2:

为了确定第一数据对应的第一向量，在上述实施例的基础上，在本申请实施例中，所述将待匹配的第一数据输入到预先训练完成的向量转化模型中，获得所述第一数据对应的第一向量包括:

确定所述第一数据对应的第一目标数据类型;

根据所述第一目标数据类型以及预先保存的数据类型和预先训练完成的向量转化模型的对应关系，确定所述第一数据对应的预先训练完成的第一目标向量转化模型;

将所述第一数据输入到所述预先训练完成的第一目标向量转化模型中，获得所述第一数据对应的所述第一向量。

在本申请实施例中，由于待匹配的第一数据可以为文本数据，比如，该第一数据可以为姓名、性别、地址等，也可以为数字数据，比如，该第一数据可以为身份证号、银行卡号以及准考证号等，也可以为图像数据，比如，该第一数据可以为用于进行人脸识别的图像等，因此，为了针对不用数据类型的第一数据，获得对应的第一向量的预先训练完成的向量转化模型也可以

不相同。

具体的，可以在第一设备中保存有数据类型和预先训练完成的向量转化模型的对应关系，根据获取到的待匹配的第一数据对应的第一目标数据类型，采用对应的预先训练完成的向量转化模型，获取该第一数据对应的第一向量；其中，该对应的预先训练完成的向量转化模型也就是预先训练完成的第一目标向量转化模型。

其中，本申请对第一数据对应的第一目标数据类型不作具体限定，示例性的，第一目标数据类型可以为文本类型、或数字类型、或图像类型等。另外，本申请对数据类型对应的向量转化模型也不作具体限定，可以根据需求灵活设置。

示例性的，为了准确的确定将第一数据转换为第一向量的向量转化模型，在上述各实施例的基础上，在本申请实施例中，若所述第一目标数据类型为文本类型，则所述第一目标数据类型对应的预先训练完成的第一目标向量转化模型为可以词向量模型或句向量模型；若所述第一目标数据类型为数字类型，则所述第一目标数据类型对应的预先训练完成的第一目标向量转化模型可以为独热（One-Hot）编码模型；若所述第一目标数据类型为图像类型，所述第一目标数据类型对应的预先训练完成的第一目标向量转化模型可以为图像向量模型。

具体的，若第一数据为文本数据，即第一数据的第一目标数据类型为文本类型，则根据预先保存的数据类型和预先训练完成的向量转化模型的对应关系，可以确定该第一目标数据类型对应的预先训练完成的第一目标向量转化模型，该第一目标向量转化模型可以为词向量模型或者句向量模型，并可以基于该预先训练完成的词向量模型或者句向量模型，获取该第一数据对应的第一向量。

在一种可能的实施方式中，若第一数据为数字数据，即第一数据的第一目标数据类型为数字类型，则根据预先保存的数据类型和预先训练完成的向量转化模型的对应关系，可以确定该第一目标数据类型对应的预先训练完成

的第一目标向量转化模型，该第一目标向量转化模型可以为预先训练完成的One-Hot 编码模型，并可以基于该预先训练完成的 One-Hot 编码模型，获取该第一数据对应的第一向量。

在一种可能的实施方式中，若第一数据为图像数据，即第一数据的第一目标数据类型为图像类型，则根据预先保存的数据类型和预先训练完成的向量转化模型的对应关系，可以确定该第一目标数据类型对应的预先训练完成的第一目标向量转化模型，该第一目标向量转化模型可以为预先训练完成的图像向量模型（图像向量嵌入模型），并可以基于该预先训练完成的图像向量模型，获取该第一数据对应的第一向量。

为方便理解，以预先训练完成的向量转化模型输出的向量的维度为 5，预先训练完成的向量转化模型为词向量模型为例进行说明。若第一数据为文本数据，且该第一数据为“某海市某东新区晴某小卖部”则将“某海市某东新区晴某小卖部”输入到预先训练完成的词向量模型中，输出的该“某海市某东新区晴某小卖部”对应的第一向量可以为（1.0，2.0，1.5，2.0，3.5）。

若预先训练完成的向量转化模型为 One-Hot 编码模型，可以预先针对每个数字设置该对应的独热编码，比如，数字包含 0-9，则该 0-9 各个数字中，0 对应的独热编码为 0000000001、1 对应的独热编码为 0000000010、2 对应的独热编码为 0000000100、3 对应的独热编码为 0000001000、4 对应的独热编码为 0000010000、5 对应的独热编码为 0000100000、6 对应的独热编码为 0001000000、7 对应的独热编码为 0010000000、8 对应的独热编码为 0100000000、9 对应的独热编码为 1000000000。将数字数据输入到 One-Hot 编码模型中，该 One-Hot 编码模型输出的第一向量中每个第一分量为该第一数据中对应的每个数字的独热编码。

示例性的，若第一数据为数字数据，该数字数据为“12345”，则将“12345”输入到预先训练完成的词向量模型中，输出的该“12345”对应的第一向量为（0000000010，0000000100，0000001000，0000010000，0000100000）。

其中，在对向量转化模型进行训练时，可以预先标注每个数据和数据对

应的标注向量，将每个数据和对应的标注向量输入到原始向量转化模型中，根据原始向量转化模型输出的预测向量及对应的标注向量，对该原始向量转化模型的参数进行调整，当满足收敛条件时，确定向量转化模型训练完成。

在本申请实施例中，无论第一数据和第二数据是数字数据还是文本数据，都能实现模糊匹配，进一步扩宽了应用场景。

实施例 3:

为了使得第二设备也能确定第一数据以及第二数据是否匹配，在上述各实施例的基础上，在本申请实施例中，所述确定所述第一向量和所述第二向量的目标距离之后，所述方法还包括:

将所述目标距离发送给所述第二设备，使所述第二设备基于所述目标距离以及预设的第一距离阈值，确定所述第一数据以及所述第二数据是否匹配。

在一种可能的实施方式中，为了使得第二设备也能确定第一数据与第二数据是否匹配，第一设备可以将确定的第一向量和第二向量的目标距离发送给第二设备，第二设备在接收到第一设备发送的第一向量和第二向量的目标距离后，可以基于该目标距离以及预设的第一距离阈值，确定第一数据与第二数据是否匹配。其中第二设备确定第一数据和第二数据是否匹配的过程与上述实施例中第一设备确定第一数据和第二数据是否匹配的过程可以相同，在此不再赘述。

在一种可能的实施方式中，第一设备确定了第一数据与第二数据是否匹配，得到第一数据与第二数据是否匹配的匹配结果之后，可以将该匹配结果发送给第二设备。可选的，第二设备可以将第一设备确定的匹配结果与自身确定的匹配结果进行比对，以进一步提高准确性。另外，在一种可能的实施方式中，第二设备也可以不进行确定第一数据与第二数据是否匹配，得到匹配结果的过程，直接利用第一设备确定的匹配结果，从而可以节省能耗。

实施例 4:

为了使得第二设备也能确定第一向量和第二向量的目标距离，在上述各实施例的基础上，所述方法还包括:

基于所述 OT 协议生成第二随机向量；基于所述第一向量、所述第二随机向量以及所述第二设定算法生成第三距离中间结果；

将所述第三距离中间结果发送给所述第二设备，使所述第二设备基于所述第二向量、所述第三距离中间结果和第四距离中间结果，确定所述第一向量和第二向量的目标距离，并使所述第二设备基于所述目标距离以及预设的第一距离阈值，确定所述第一数据以及所述第二数据是否匹配；其中，所述第四距离中间结果为基于第二中间向量以及所述第一设定算法获得的，所述第二中间向量为通过所述 OT 协议获取的，且所述第二中间向量是基于所述第二向量对应的二进制向量、第一向量及所述第二随机向量生成的。

在一种可能的实施方式中，为了使得第二设备也能确定第一向量和第二向量的目标距离，而无需依赖第一设备将其确定的目标距离发送给第二设备，进一步提高安全性，第一设备也可以基于 OT 协议生成随机向量（随机数，为方便描述，称为第二随机向量）。第一设备也可以基于第一向量、第二随机向量以及第二设定算法，生成第三距离中间结果。第一设备生成第三距离中间结果的过程与第二设备生成第二距离中间结果的过程类似，在一种可能的实施方式中，第三距离中间结果可以是采用以下过程获得的：

针对第一向量中的每个第一分量x_i，确定该第一分量的第一平方分量(x_i^2)；并可以基于第二随机向量中相应的第二随机分量C_j'、及该第二随机分量在相应的二进制向量中的位置信息（如 p-j 等），确定该第一分量对应的第三随机分量子值（如$\sum_{j=0}^{p} 2^{p-j+1} C_j'$）；并可以基于该第一平方分量与该第三随机分量子值的和，确定第三距离中间结果。

为方便理解，下面用公式形式对基于第一向量、第二随机向量以及第二设定算法生成第三距离中间结果的过程进行解释说明。假设第三距离中间结果用 E(a)′表示，则：$E(a)' = \sum_{i=1}^{m}\left(x_i^2 + \sum_{j=0}^{p} 2^{p-j+1} C_j\right)$。

在一种可能的实施方式中，为了保证双方数据不完全相同时也能实现模糊匹配以及保证匹配过程的安全性，第二设备获得第二数据对应的第二向量

之后，可以继而将第二向量转化为二进制的二进制向量。示例性的，假设第二数据 U5 对应的第二向量为（y_1，y，y_3……，y_m），第二向量中包含 m 个第二分量(y_i)，其中，i 为不大于 m 的任一正整数，可以将每个十进制的第一分量(y_i)分别转化为二进制的二进制向量（二进制的 0-1 序列），即将$(y_i)_{10}$转化为（$b_0b_1b_2 \ldots b_j \ldots b_p$）$_2$，$(y_i)_{10}$ =（$b_0b_1b_2 \ldots b_j \ldots b_p$）$_2$，其中$(y_i)_{10}$表示十进制的第二分量，（$b_0b_1b_2 \ldots b_j \ldots b_p$）$_2$表示该第二分量对应的二进制的二进制向量，每个二进制向量中包含 p 个二进制分量b_j，其中 j 为不大于 p 的任一正整数，b ∈（0,1），则$y_i = \sum_{j=0}^{p} 2^{p-j} b_j$。将第二向量中的每个第二分量均转化为相应的二进制向量后，即得到第二向量对应的二进制向量，第二向量对应的二进制向量可以认为是由每个第二分量对应的二进制向量组成的二进制向量集。

第二设备可以通过 OT 协议，获取第二中间向量，第二中间向量中包含的每个第二中间分量可以为：$b_jx_i + C_j'$。其中，OT 协议可以为 1-out-2 OT 协议，第二设备可以作为 OT 协议的输入方，第一设备可以作为 OT 协议的输出方。C_j'为第一设备基于 OT 协议生成的随机向量(随机数)中包含的第二随机分量。可以理解的，当$b_j = 0$时，则第二设备获取的第二中间分量$b_jx_i + C_j'$为C_j'；当$b_j = 1$时，则第一设备获取到的第一中间分量$b_jx_i + C_j'$为$x_i + C_j'$。

获得第二中间向量之后，第二设备可以基于第二中间向量以及第一设定算法，获得第四距离中间结果。第二设备获得第四距离中间结果的过程与上述实施例中第一设备获得第一距离中间结果的过程类似，具体的，获得第四距离中间结果可以是采用以下过程获得的：

针对第二向量中的每个第二分量y_i，可以基于该第二分量对应的每个二进制分量b_j及第一向量中相应的第一分量x_i，确定该第二分量与相应的第一分量的乘积(x_i y_i)；并可以基于第二随机向量中相应的第二随机分量C_j'、及该第二随机分量在相应的二进制向量中的位置信息（如 p-j 等），确定该第二分量对应的第四随机分量子值（如$-2\sum_{j=0}^{p} 2^{p-j} C_j'$）。可以基于该第二分量与相应的

第一分量的乘积($x_i\ y_i$)与该第四随机分量子值的和，确定第四距离中间结果。

为方便理解，下面用公式形式对基于第二中间向量以及第一设定算法，确定第四距离中间结果的过程进行解释说明。假设第四距离中间结果用 $E(b)'$ 表示，则 $E(b)'=\sum_{i=1}^{m}\{-2*[\sum_{j=0}^{p}2^{p-j}(b_jx_i+C_j')]\}$

$=\sum_{i=1}^{m}\{-2*[\sum_{j=0}^{p}2^{p-j}b_jx_i+\sum_{j=0}^{p}2^{p-j}C_j']\}$

$=\sum_{i=1}^{m}\{-2*[x_iy_i+\sum_{j=0}^{p}2^{p-j}C_j']\}$

$=\sum_{i=1}^{m}(-2x_iy_i-2\sum_{j=0}^{p}2^{p-j}C_j')$。

在一种可能的实施方式中，第一设备可以将确定的第三距离中间结果 $E(a)'$ 发送给第二设备。第二设备可以进而基于第二向量、第三距离中间结果和第四距离中间结果，确定第一向量和第二向量的目标距离。可选的，基于第二向量、第三距离中间结果和第四距离中间结果，确定第一向量和第二向量的目标距离的过程可以如下：

第二设备针对第二向量中的每个第二分量，可以先基于该第二分量的第二平方分量 y_i^2 与第三距离中间结果 $E(a)'$ 和第四距离中间结果 $E(b)'$ 的和（$y_i^2+E(a)'+E(b)'$），确定该第二分量对应的子距离 $(x_i-y_i)^2$，然后基于每个第二分量对应的子距离的和，确定第一向量和第二向量的目标距离。

例如，计算每个第二分量对应的子距离的和的过程可以如下：

$\sum_{i=1}^{m}(y_i^2)+E(a)'+E(b)'$

$=\sum_{i=1}^{m}(y_i^2)+\sum_{i=1}^{m}(-2x_iy_i-2\sum_{j=0}^{p}2^{p-j}C_j')+\sum_{i=1}^{m}(x_i^2+\sum_{j=0}^{p}2^{p-j+1}C_j')$

$=\sum_{i=1}^{m}(x_i^2)-2\sum_{i=1}^{m}(x_iy_i)+\sum_{i=1}^{m}(y_i^2)$

$=\sum_{i=1}^{m}(x_i-y_i)^2$。

可以基于每个第二分量对应的子距离（$\sum_{i=1}^{m}(x_i-y_i)^2$）以及欧氏距离公式，计算第一向量和第二向量之间的目标距离，示例性的，第一向量和第二向量之间的目标距离 $d=\sqrt{\sum_{i=1}^{m}(x_i-y_i)^2}$。确定了第一向量和第二向量的目标距离之后，也可以基于该目标距离以及预设的第一距离阈值，确定第一数据和第二数据是否匹配，其中确定第一数据和第二数据是否匹配的过程与上述实施

例中第一设备确定第一数据和第二数据是否匹配的过程可以相同，在此不再赘述。

为方便理解，下面通过一个具体实施例对本申请提供的数据匹配过程进行说明。图 2 为本申请实施例提供的第二种数据匹配过程示意图，如图 2 所示，该过程包括以下操作:

第一设备分别将待匹配的每个第一数据输入到部署在第一设备中的预先训练完成的向量转化模型中，针对每个第一数据，获得该第一数据对应的第一向量。同样的，第二设备分别将待匹配的每个第二数据输入到部署在第二设备中的预先训练完成的向量转化模型中，针对每个第二数据，获得该第二数据对应的第二向量。如图 2 所述，分别有 4 个第一数据，分别为 U1、U2、U3、U4，且 U1 对应的第一向量为（x11，x12，x13……，x1m），U2 对应的第一向量为（x21，x22，x23……，x2m），U3 对应的第一向量为（x31，x32，x33……，x3m），U4 对应的第一向量为（x41，x42，x43……，x4m）。另外，分别有 4 个第二数据，分别为 U5、U6、U7、U8，且 U5 对应的第二向量为（y11，y12，y13……，y1m），U6 对应的第二向量为（y21，y22，y23……，y2m），U7 对应的第二向量为（y31，y32，y33……，y3m），U8 对应的第二向量为（y41，y42，y43……，y4m）。

针对每个第一向量，第一设备将该第一向量转化为二进制的二进制向量，获得该第一向量对应的二进制向量。第二设备基于 OT 协议生成第一随机向量。

由于针对每个第一数据，需要确定该第一数据与每个第二数据是否匹配，例如，针对第一数据 U1，需要确定第一数据 U1 与第二数据 U5 是否匹配、第一数据 U1 与第二数据 U6 是否匹配、第一数据 U1 与第二数据 U7 是否匹配、第一数据 U1 与第二数据 U8 是否匹配，因此，针对每个第一向量，均可以确定该第一向量与每个第二向量的目标距离，其中，确定任一第一向量与任一第二向量的目标距离的过程均与上述实施例中计算目标距离的过程相同，例如，针对每个第一向量，第一设备通过 OT 协议，获取第一中间向量，基于该第一中间向量以及第一设定算法，获得第一距离中间结果。其中，第一中间

向量是基于第一向量对应的二进制向量、第二向量及第一随机向量生成的。第二设备基于第二向量、第一随机向量以及第二设定算法生成的第二距离中间结果，并将第二距离中间结果发送给第一设备。针对每个第一向量，第一设备基于该第一向量、第一距离中间结果和第二距离中间结果，确定该第一向量和第二向量的目标距离，在此不再赘述。

针对每个第一向量，分别确定了该第一向量与每个第二向量的目标距离之后，第一设备还可以将确定的目标距离发送给第二设备，第一设备和第二设备均可以分别基于该目标距离以及预设的第一距离阈值，确定该第一向量对应的第一数据分别与每个第二向量对应的第二数据是否匹配，即可以分别确定各个第一数据与各个第二数据是否匹配。

为方便理解，下面再通过一个具体实施例对本申请提供的数据匹配过程进行说明。假设保存在第一设备中的第一数据有 3 个，分别为“某海市某东新区晴某小卖部”、“某海市天某餐饮店”、“高某路杨某福麻辣烫”。保存在第二设备中的第二数据也有 3 个，分别为“某海市晴某小卖部”、“某海市天某餐饮店”、“麦某劳”。

其中第一数据中，“某海市某东新区晴某小卖部”对应的第一向量为 $<1.0,2.0,1.5,2.0,3.5>$，记为 A1。“某海市天某餐饮店”对应的第一向量为 $<3.0,4.0,2.5,2.5,1.5>$，记为 A2。“高某路杨某福麻辣烫”对应的第一向量为 $<4.5,5.5,7.5,1.5,0.5>$，记为 A3。

第二数据中，“某海市晴某小卖部”对应的第二向量为 $<1.0,2.0,1.5,1.0,3.5>$，记为 B1。“某海市天某餐饮店”对应的第二向量为 $<3.0,4.0,2.5,2.5,1.5>$，记为 B2。“麦某劳”对应的第二向量为 $<3.5,6.5,2.5,7.5,2.5>$，记为 B3。

由于针对每个第一数据，需要确定该第一数据与每个第二数据是否匹配，因此，针对每个第一向量，均可以确定该第一向量与每个第二向量的目标距离，其中，确定任一第一向量与任一第二向量的目标距离的过程均与上述实施例中计算目标距离的过程相同，在此不再赘述。

如表 1 所示，表 1 为本申请一些实施例提供的一种目标距离的示意表。

表 1

D(x,y)	B1	B2	B3
A1	1	3.64	7.66
A2	3.9	0	5.7
A3	8.35	5.61	8.18

假设 D(x,y)表征目标距离，其中 x 表征的第一向量，y 表征的第二向量。A1 与 B1 对应的目标距离为 1，A1 与 B2 对应的目标距离为 3.64，A1 与 B3 对应的目标距离为 7.66，A2 与 B1 对应的目标距离为 3.9，A2 与 B2 对应的目标距离为 0，A2 与 B3 对应的目标距离为 5.7，A3 与 B1 对应的目标距离为 8.35，A3 与 B2 对应的目标距离为 5.16，A3 与 B3 对应的目标距离为 8.18。

假设第一距离阈值为 2，则本次模糊匹配结果为第一数据中的“某海市某东新区晴某小卖部”与第二数据中的“某海市晴某小卖部”匹配，第一数据中的“某海市天某餐饮店”与第二数据中的“某海市天某餐饮店”匹配，由此也可以看出，本申请中的数据匹配方法可以使得两个数据在不完全相同时，也能实现数据的模糊匹配，拓宽了使用场景。

为方便理解，下面再通过一个具体实施例对本申请提供的数据匹配过程进行说明。假设保存在第一设备中的第一数据有 3 个，分别为 3 个手机号码“13345678909”、“13245678911”、“13334536787”。保存在第二设备中的第二数据也有 3 个，分别为 3 个手机号码“13334536787”、“13345678908”、“15439402290”。

第一设备和第二设备分别可以利用独热（One-Hot）编码模型（数据转化模型）第一数据和第二数据中的手机号码进行向量化，生成相应的第一向量和第二向量，可选的，第一向量和第二向量的向量维度可以为 10*11 维，其中 10 代表 0-9 共 10 个数字，11 代表手机号长度为 11 个数字。示例性的，参阅表 2，表 2 为手机号码“13345678909”对应的向量示意表。

表 2

手机号码中的其中一个数字	数字对应的向量（独热编码）
1	0000000010
3	0000001000
3	0000001000
4	0000010000
5	0000100000
6	0001000000
7	0010000000
8	0100000000
9	1000000000
0	0000000001
9	1000000000

参阅表 2，在一种可能的实施方式中，0 对应的独热编码为 0000000001、1 对应的独热编码为 0000000010、2 对应的独热编码为 0000000100、3 对应的独热编码为 0000001000、4 对应的独热编码为 0000010000、5 对应的独热编码为 0000100000、6 对应的独热编码为 0001000000、7 对应的独热编码为 0010000000、8 对应的独热编码为 0100000000、9 对应的独热编码为 1000000000。则手机号码“13345678909”对应的第一向量（或第二向量）可以为（0000000010，0000001000，0000001000，0000010000，0000100000，0001000000，0010000000，0100000000，1000000000，0000000001，1000000000）。确定其他手机号码对应的向量的过程与该过程类似，在此不再赘述。

为方便描述，将第一数据中“13345678909”对应的第一向量记为 A4、“13245678911”对应的第一向量记为 A5、“13334536787”对应的第一向量记为 A6。将第二数据中“13334536787”对应的第二向量记为 B4、“13345678908”对应的第二向量记为 B5、“15439402290”对应的第二向量记为 B6。

由于针对每个第一数据，需要确定该第一数据与每个第二数据是否匹配，因此，针对每个第一向量，均可以确定该第一向量与每个第二向量的目标距离，其中，确定任一第一向量与任一第二向量的目标距离的过程均与上述实施例中计算目标距离的过程相同，在此不再赘述。

如表 3 所示，表 3 为本申请一些实施例提供的一种目标距离的示意表。

表 3

D(x,y)	B4	B5	B6
A4	2.82	1	3.16
A5	3	1.73	3
A6	0	2.82	3

假设 D(x,y)表征目标距离，其中 x 表征的第一向量，y 表征的第二向量。A4 与 B4 对应的目标距离为 2.82，A4 与 B5 对应的目标距离为 1，A4 与 B6 对应的目标距离为 3.16，A5 与 B4 对应的目标距离为 3，A5 与 B5 对应的目标距离为 1.73，A5 与 B6 对应的目标距离为 3，A6 与 B4 对应的目标距离为 0，A6 与 B5 对应的目标距离为 2.83，A6 与 B6 对应的目标距离为 3。

假设第一距离阈值为 2，则本次模糊匹配结果为第一数据中的“13345678909”与第二数据中的“13345678909”匹配，第一数据中的“13334536787”与第二数据中的“13334536787”匹配，由此也可以看出，本申请中的数据匹配方法不仅能实现数据的模糊匹配，还可以实现数据的精准匹配，拓宽了使用场景。

实施例 5:

基于相同的技术构思，本申请提供了一种数据匹配方法，应用于第二设备，图 3 为本申请实施例提供的第三种数据匹配过程示意图，如图 3 所示，该方法包括:

S301：将待匹配的第二数据输入到预先训练完成的向量转化模型中，获得所述第二数据对应的第二向量；并基于不经意传输 OT 协议，生成第一随机向量。

本申请实施例提供的数据匹配方法应用于第二设备，该第二设备可以是智能终端、PC 或者服务器等设备，且该第二设备与本申请中的第一设备为不同的设备。

S302：基于所述第二向量、所述第一随机向量以及第二设定算法生成的第二距离中间结果，并将所述第二距离中间结果发送给第一设备，使第一设

备基于第一向量、第一距离中间结果和所述第二距离中间结果，确定所述第一向量和第二向量的目标距离；并使所述第一设备基于所述目标距离以及预设的第一距离阈值，确定所述第一数据以及所述第二数据是否匹配；其中，所述第一向量为将第一数据输入到所述第一设备中的预先训练完成的向量转化模型中获得的；所述第一距离中间结果为基于第一中间向量以及第一设定算法获得的，所述第一中间向量为基于所述第一向量对应的二进制向量、所述第二向量及所述第一随机向量生成的。

在一种可能的实施方式中，所述将待匹配的第二数据输入到预先训练完成的向量转化模型中，获得所述第二数据对应的第二向量包括:

确定所述第二数据对应的第二目标数据类型;

根据所述第二目标数据类型以及预先保存的数据类型和预先训练完成的向量转化模型的对应关系，确定所述第二数据对应的预先训练完成的第二目标向量转化模型;

将所述第二数据输入到所述预先训练完成的第二目标向量转化模型中，获得所述第二数据对应的所述第二向量。

在一种可能的实施方式中，所述第二目标数据类型为文本类型、数字类型、图像类型中的任意一种。

在一种可能的实施方式中，若所述第二目标数据类型为文本类型，所述第二目标数据类型对应的预先训练完成的第二目标向量转化模型为词向量模型或句向量模型；若所述第二目标数据类型为数字类型，所述第二目标数据类型对应的预先训练完成的第二目标向量转化模型为独热编码模型；若所述第二目标数据类型为图像类型，所述第二目标数据类型对应的预先训练完成的第二目标向量转化模型为图像向量模型。

在一种可能的实施方式中，所述方法还包括:

接收所述第一设备发送的第三距离中间结果，所述第三距离中间结果为基于第一向量、第二随机向量以及所述第二设定算法生成的，其中，所述第二随机向量为所述第一设备基于所述 OT 协议生成的；并通过 OT 协议获取第

二中间向量，其中所述第二中间向量是基于所述第二向量对应的二进制向量、第一向量及所述第二随机向量生成的；

基于所述第二中间向量以及所述第一设定算法获取第四距离中间结果；

基于所述第二向量、所述第三距离中间结果和第四距离中间结果，确定所述第一向量和第二向量的目标距离；

基于所述目标距离以及预设的第一距离阈值，确定所述第一数据以及所述第二数据是否匹配。

在一种可能的实施方式中，所述将所述第二距离中间结果发送给所述第一设备之后，所述方法还包括：

接收所述第一设备发送的目标距离，基于所述目标距离以及预设的第一距离阈值，确定所述第一数据以及所述第二数据是否匹配。

在一种可能的实施方式中，所述方法还包括：

接收所述第一设备发送的所述第一数据以及所述第二数据是否匹配的匹配结果。

实施例 6：

基于相同的技术构思，本申请提供了一种数据匹配装置，应用于第一设备。图 4 为本申请一些实施例提供的一种数据匹配装置结构示意图，如图 4 所示，所述装置包括：

第一获取模块 41，用于将待匹配的第一数据输入到预先训练完成的向量转化模型中，获得所述第一数据对应的第一向量，并获得所述第一向量对应的二进制向量；

第一处理模块 42，用于通过不经意传输 OT 协议，获取第一中间向量；基于所述第一中间向量以及第一设定算法，获得第一距离中间结果；其中，所述第一中间向量是基于所述第一向量对应的二进制向量、第二向量及第一随机向量生成的，所述第二向量为第二设备将第二数据输入到所述第二设备中的预先训练完成的向量转化模型中获得的，所述第一随机向量为所述第二设备基于所述 OT 协议生成的；

所述第一处理模块 42，还用于接收所述第二设备发送的基于所述第二向量、所述第一随机向量以及第二设定算法生成的第二距离中间结果；基于所述第一向量、所述第一距离中间结果和所述第二距离中间结果，确定所述第一向量和第二向量的目标距离；

第一确定模块 43，用于基于所述目标距离以及预设的第一距离阈值，确定所述第一数据以及所述第二数据是否匹配。

在一种可能的实施方式中，所述第一获取模块 41，具体用于确定所述第一数据对应的第一目标数据类型；

根据所述第一目标数据类型以及预先保存的数据类型和预先训练完成的向量转化模型的对应关系，确定所述第一数据对应的预先训练完成的第一目标向量转化模型；

将所述第一数据输入到所述预先训练完成的第一目标向量转化模型中，获得所述第一数据对应的所述第一向量。

在一种可能的实施方式中，所述第一目标数据类型为文本类型、数字类型、图像类型中的至少一种。

在一种可能的实施方式中，若所述第一目标数据类型为文本类型，所述第一目标数据类型对应的预先训练完成的第一目标向量转化模型为词向量模型或句向量模型；若所述第一目标数据类型为数字类型，所述第一目标数据类型对应的预先训练完成的第一目标向量转化模型为独热编码模型；若所述第二目标数据类型为图像类型，所述第二目标数据类型对应的预先训练完成的第二目标向量转化模型为图像向量模型。

在一种可能的实施方式中，所述第一处理模块 42，还用于基于所述 OT 协议生成第二随机向量；基于所述第一向量、所述第二随机向量以及所述第二设定算法生成第三距离中间结果；将所述第三距离中间结果发送给所述第二设备，使所述第二设备基于所述第二向量、所述第三距离中间结果和第四距离中间结果，确定所述第一向量和第二向量的目标距离，并使所述第二设备基于所述目标距离以及预设的第一距离阈值，确定所述第一数据以及所述

第二数据是否匹配；其中，所述第四距离中间结果为基于第二中间向量以及所述第一设定算法获得的，所述第二中间向量为通过所述 OT 协议获取的，且所述第二中间向量是基于所述第二向量对应的二进制向量、第一向量及所述第二随机向量生成的。

在一种可能的实施方式中，所述第一处理模块 42，还用于将所述目标距离发送给所述第二设备，使所述第二设备基于所述目标距离以及预设的第一距离阈值，确定所述第一数据以及所述第二数据是否匹配。

在一种可能的实施方式中，所述第一处理模块 42，还用于将确定的所述第一数据以及所述第二数据是否匹配的匹配结果发送给所述第二设备。

在一种可能的实施方式中，所述第一向量和所述第二向量的长度均为预设长度。

在一种可能的实施方式中，所述第一距离中间结果是采用以下过程获得的：

针对所述第一向量中的每个第一分量，基于该第一分量对应的每个二进制分量及所述第二向量中相应的第二分量，确定该第一分量与相应的第二分量的乘积；并基于所述第一随机向量中相应的第一随机分量、及该第一随机分量在相应的二进制向量中的位置信息，确定该第一分量对应的第一随机分量子值；基于该乘积与该第一随机分量子值的和，确定所述第一距离中间结果；

所述第二距离中间结果是采用以下过程获得的：

针对所述第二向量中的每个第二分量，确定该第二分量的第二平方分量；并基于所述第一随机向量中相应的第一随机分量、及该第一随机分量在相应的二进制向量中的位置信息，确定该第二分量对应的第二随机分量子值；基于该第二平方分量与该第二随机分量子值的和，确定所述第二距离中间结果；

基于所述第一向量、所述第一距离中间结果和所述第二距离中间结果，确定所述第一向量和第二向量的目标距离，包括：

针对所述第一向量中的每个第一分量，确定该第一分量的第一平方分量；

基于该第一分量的第一平方分量、该第一分量与相应的第二分量的乘积、第一随机分量子值、相应的第二分量的第二平方分量以及第二随机分量子值，确定该第一分量对应的子距离；

基于每个第一分量对应的子距离，确定所述第一向量和第二向量的目标距离。

在一种可能的实施方式中，所述第三距离中间结果是采用以下过程获得的：

针对所述第一向量中的每个第一分量，确定该第一分量的第一平方分量；并基于所述第二随机向量中相应的第二随机分量，及该第二随机分量在相应的二进制向量中的位置信息，确定该第一分量对应的第三随机分量子值；基于该第一平方分量与该第三随机分量子值的和，确定所述第三距离中间结果；

所述第四距离中间结果是采用以下过程获得的：

针对所述第二向量中的每个第二分量，基于该第二分量对应的每个二进制分量及所述第一向量中相应的第一分量，确定该第二分量与相应的第一分量的乘积；并基于所述第二随机向量中相应的第二随机分量、及该第二随机分量在相应的二进制向量中的位置信息，确定该第二分量对应的第四随机分量子值；基于该乘积与该第四随机分量子值的和，确定所述第四距离中间结果；

基于所述第二向量、所述第三距离中间结果和所述第四距离中间结果，确定所述第一向量和第二向量的目标距离，包括：

针对所述第二向量中的每个第二分量，确定该第二分量的第二平方分量；基于该第二分量的第二平方分量、该第二分量与相应的第一分量的乘积、第三随机分量子值、相应的第一分量的第一平方分量以及第四随机分量子值，确定该第二分量对应的子距离；

基于每个第二分量对应的子距离，确定所述第一向量和第二向量的目标距离。

在一种可能的实施方式中，所述第一确定模块 43，具体用于判断所述目

标距离是否小于预设的第一距离阈值;

若是，则确定所述第一数据与所述第二数据匹配;

否则，确定所述第一数据与所述第二数据不匹配。

在一种可能的实施方式中，所述第一确定模块 43，还用于判断所述目标距离是否等于预设的第二距离阈值，若是，则确定所述第一数据与所述第二数据相同。

实施例 7:

基于相同的技术构思，本申请提供了另一种数据匹配装置，应用于第二设备。图 5 为本申请一些实施例提供的另一种数据匹配装置结构示意图，如图 5 所示，所述装置包括:

第二获取模块 51，用于将待匹配的第二数据输入到预先训练完成的向量转化模型中，获得所述第二数据对应的第二向量;并基于不经意传输 OT 协议，生成第一随机向量;

第二处理模块 52，用于基于所述第二向量、所述第一随机向量以及第二设定算法生成的第二距离中间结果，并将所述第二距离中间结果发送给第一设备，使第一设备基于第一向量、第一距离中间结果和所述第二距离中间结果，确定所述第一向量和第二向量的目标距离;并使所述第一设备基于所述目标距离以及预设的第一距离阈值，确定所述第一数据以及所述第二数据是否匹配;其中，所述第一向量为将第一数据输入到所述第一设备中的预先训练完成的向量转化模型中获得的;所述第一距离中间结果为基于第一中间向量以及第一设定算法获得的，所述第一中间向量为基于所述第一向量对应的二进制向量、所述第二向量及所述第一随机向量生成的。

在一种可能的实施方式中，所述第二获取模块 51，具体用于确定所述第二数据对应的第二目标数据类型;

根据所述第二目标数据类型以及预先保存的数据类型和预先训练完成的向量转化模型的对应关系，确定所述第二数据对应的预先训练完成的第二目标向量转化模型;

将所述第二数据输入到所述预先训练完成的第二目标向量转化模型中，获得所述第二数据对应的所述第二向量。

在一种可能的实施方式中，所述第二目标数据类型为文本类型、数字类型、图像类型中的至少一种。

在一种可能的实施方式中，若所述第二目标数据类型为文本类型，所述第二目标数据类型对应的预先训练完成的第二目标向量转化模型为词向量模型或句向量模型；若所述第二目标数据类型为数字类型，所述第二目标数据类型对应的预先训练完成的第二目标向量转化模型为独热编码模型；若所述第二目标数据类型为图像类型，所述第二目标数据类型对应的预先训练完成的第二目标向量转化模型为图像向量模型。

在一种可能的实施方式中，所述第二处理模块 52，还用于接收所述第一设备发送的第三距离中间结果，所述第三距离中间结果为基于第一向量、第二随机向量以及所述第二设定算法生成的，其中，所述第二随机向量为所述第一设备基于所述 OT 协议生成的；并通过 OT 协议获取第二中间向量，其中所述第二中间向量是基于所述第二向量对应的二进制向量、第一向量及所述第二随机向量生成的；

基于所述第二中间向量以及所述第一设定算法获取第四距离中间结果；

基于所述第二向量、所述第三距离中间结果和第四距离中间结果，确定所述第一向量和第二向量的目标距离；

基于所述目标距离以及预设的第一距离阈值，确定所述第一数据以及所述第二数据是否匹配。

在一种可能的实施方式中，所述第二处理模块 52，还用于接收所述第一设备发送的目标距离，基于所述目标距离以及预设的第一距离阈值，确定所述第一数据以及所述第二数据是否匹配。

在一种可能的实施方式中，所述第二处理模块 52，还用于接收所述第一设备发送的所述第一数据以及所述第二数据是否匹配的匹配结果。

实施例 8:

基于相同的技术构思，本申请提供了一种数据匹配系统，图 6 为本申请一些实施例提供的一种数据匹配系统结构示意图，如图 6 所示，所述系统包括：

第一设备 61，用于将待匹配的第一数据输入到预先训练完成的向量转化模型中，获得所述第一数据对应的第一向量，并获得所述第一向量对应的二进制向量；

第二设备 62，用于将待匹配的第二数据输入到预先训练完成的向量转化模型中，获得所述第二数据对应的第二向量；将待匹配的第二数据输入到预先训练完成的向量转化模型中，获得所述第二数据对应的第二向量；并基于不经意传输 OT 协议，生成第一随机向量；

所述第一设备 61，还用于通过不经意传输 OT 协议，获取第一中间向量；基于所述第一中间向量以及第一设定算法，获得第一距离中间结果；其中，所述第一中间向量是基于所述第一向量对应的二进制向量、第二向量及第一随机向量生成的；

所述第二设备 62，还用于基于所述第二向量、所述第一随机向量以及第二设定算法生成的第二距离中间结果，并将所述第二距离中间结果发送给所述第一设备；

所述第一设备 61，还用于基于所述第一向量、所述第一距离中间结果和所述第二距离中间结果，确定所述第一向量和第二向量的目标距离；基于所述目标距离以及预设的第一距离阈值，确定所述第一数据以及所述第二数据是否匹配。

基于相同的技术构思，本申请还提供了一种电子设备，图 7 为本申请一些实施例提供的一种电子设备结构示意图，如图 7 所示，包括：处理器 71、通信接口 72、存储器 73 和通信总线 74，其中，处理器 71，通信接口 72，存储器 73 通过通信总线 74 完成相互间的通信；

所述存储器 73 中存储有计算机程序，当所述程序被所述处理器 71 执行时，使得所述处理器 71 执行上述任一实施例的数据匹配方法的步骤。

由于上述电子设备解决问题的原理与数据匹配方法相似，因此上述电子设备的实施可以参见方法的实施，重复之处不再赘述。

上述电子设备提到的通信总线可以是外设部件互连标准（Peripheral Component Interconnect，PCI）总线或扩展工业标准结构（Extended Industry Standard Architecture，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。为便于表示，图中仅用一条粗线表示，但并不表示仅有一根总线或一种类型的总线。

通信接口 72 用于上述电子设备与其他设备之间的通信。

存储器可以包括随机存取存储器（Random Access Memory，RAM），也可以包括非易失性存储器（Non-Volatile Memory，NVM），例如至少一个磁盘存储器。可选地，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。

上述处理器可以是通用处理器，包括中央处理器、网络处理器（Network Processor，NP）等；还可以是数字指令处理器（Digital Signal Processing，DSP）、专用集成电路、现场可编程门陈列或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。

基于相同的技术构思，本申请实施例提供了一种计算机可读存储介质，所述计算机可读存储介质内存储有可由电子设备执行的计算机程序，当所述程序在所述电子设备上运行时，使得所述电子设备执行时实现上述任一实施例的数据匹配方法的步骤。由于上述计算机可读存储介质解决问题的原理与数据匹配方法相似，因此上述计算机可读存储介质的实施可以参见方法的实施，重复之处不再赘述。

上述计算机可读存储介质可以是电子设备中的处理器能够存取的任何可用介质或数据存储设备，包括但不限于磁性存储器如软盘、硬盘、磁带、磁光盘（MO）等、光学存储器如 CD、DVD、BD、HVD 等、以及半导体存储器如 ROM、EPROM、EEPROM、非易失性存储器（NAND FLASH）、固态硬盘（SSD）等。

基于相同的技术构思，在上述各实施例的基础上，本申请提供了一种计

算机程序产品，所述计算机程序产品包括：计算机程序代码，当所述计算机程序代码在计算机上运行时，使得计算机执行如上述任一所述数据匹配方法的步骤。

本领域内的技术人员应明白，本申请的实施例可提供为方法、系统、或计算机程序产品。因此，本申请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，本申请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。

本申请是参照根据本申请的方法、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。

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。

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。

显然，本领域的技术人员可以对本申请进行各种改动和变型而不脱离本申请的精神和范围。这样，倘若本申请的这些修改和变型属于本申请权利要求及其等同技术的范围之内，则本申请也意图包含这些改动和变型在内。

权 利 要 求

1、一种数据匹配方法，应用于第一设备，所述方法包括:

将待匹配的第一数据输入到预先训练完成的向量转化模型中，获得所述第一数据对应的第一向量，并获得所述第一向量对应的二进制向量;

通过不经意传输 OT 协议，获取第一中间向量；基于所述第一中间向量以及第一设定算法，获得第一距离中间结果；其中，所述第一中间向量是基于所述第一向量对应的二进制向量、第二向量及第一随机向量生成的，所述第二向量为第二设备将第二数据输入到所述第二设备中的预先训练完成的向量转化模型中获得的，所述第一随机向量为所述第二设备基于所述 OT 协议生成的;

接收所述第二设备发送的基于所述第二向量、所述第一随机向量以及第二设定算法生成的第二距离中间结果；基于所述第一向量、所述第一距离中间结果和所述第二距离中间结果，确定所述第一向量和第二向量的目标距离;

基于所述目标距离以及预设的第一距离阈值，确定所述第一数据以及所述第二数据是否匹配。

2、根据权利要求 1 所述的方法，其中，所述将待匹配的第一数据输入到预先训练完成的向量转化模型中，获得所述第一数据对应的第一向量包括:

确定所述第一数据对应的第一目标数据类型;

根据所述第一目标数据类型以及预先保存的数据类型和预先训练完成的向量转化模型的对应关系，确定所述第一数据对应的预先训练完成的第一目标向量转化模型;

将所述第一数据输入到所述预先训练完成的第一目标向量转化模型中，获得所述第一数据对应的所述第一向量。

3、根据权利要求 2 所述的方法，其中，所述第一目标数据类型为文本类型、数字类型、图像类型中的任意一种。

4、根据权利要求 3 所述的方法，其中，若所述第一目标数据类型为文本

类型，所述第一目标数据类型对应的预先训练完成的第一目标向量转化模型为词向量模型或句向量模型；若所述第一目标数据类型为数字类型，所述第一目标数据类型对应的预先训练完成的第一目标向量转化模型为独热编码模型；若所述第一目标数据类型为图像类型，所述第一目标数据类型对应的预先训练完成的第一目标向量转化模型为图像向量模型。

5、根据权利要求 1 所述的方法，所述方法还包括:

基于所述 OT 协议生成第二随机向量；基于所述第一向量、所述第二随机向量以及所述第二设定算法生成第三距离中间结果;

将所述第三距离中间结果发送给所述第二设备，使所述第二设备基于所述第二向量、所述第三距离中间结果和第四距离中间结果，确定所述第一向量和第二向量的目标距离，并使所述第二设备基于所述目标距离以及预设的第一距离阈值，确定所述第一数据以及所述第二数据是否匹配；其中，所述第四距离中间结果为基于第二中间向量以及所述第一设定算法获得的，所述第二中间向量为通过所述 OT 协议获取的，且所述第二中间向量是基于所述第二向量对应的二进制向量、第一向量及所述第二随机向量生成的。

6、根据权利要求 1 所述的方法，所述确定所述第一向量和第二向量的目标距离之后，所述方法还包括:

将所述目标距离发送给所述第二设备，使所述第二设备基于所述目标距离以及预设的第一距离阈值，确定所述第一数据以及所述第二数据是否匹配。

7、根据权利要求 1 所述的方法，所述方法还包括:

将确定的所述第一数据以及所述第二数据是否匹配的匹配结果发送给所述第二设备。

8、根据权利要求 1 所述的方法，其中，所述第一向量和所述第二向量的长度均为预设长度。

9、根据权利要求 1 所述的方法，其中，所述第一距离中间结果是采用以下过程获得的:

针对所述第一向量中的每个第一分量，基于该第一分量对应的每个二进

制分量及所述第二向量中相应的第二分量，确定该第一分量与相应的第二分量的乘积；并基于所述第一随机向量中相应的第一随机分量、及该第一随机分量在相应的二进制向量中的位置信息，确定该第一分量对应的第一随机分量子值；基于该乘积与该第一随机分量子值的和，确定所述第一距离中间结果；

所述第二距离中间结果是采用以下过程获得的：

针对所述第二向量中的每个第二分量，确定该第二分量的第二平方分量；并基于所述第一随机向量中相应的第一随机分量、及该第一随机分量在相应的二进制向量中的位置信息，确定该第二分量对应的第二随机分量子值；基于该第二平方分量与该第二随机分量子值的和，确定所述第二距离中间结果；

基于所述第一向量、所述第一距离中间结果和所述第二距离中间结果，确定所述第一向量和第二向量的目标距离，包括：

针对所述第一向量中的每个第一分量，确定该第一分量的第一平方分量；基于该第一分量的第一平方分量、该第一分量与相应的第二分量的乘积、第一随机分量子值、相应的第二分量的第二平方分量以及第二随机分量子值，确定该第一分量对应的子距离；

基于每个第一分量对应的子距离，确定所述第一向量和第二向量的目标距离。

10、根据权利要求 5 所述的方法，其中，所述第三距离中间结果是采用以下过程获得的：

针对所述第一向量中的每个第一分量，确定该第一分量的第一平方分量；并基于所述第二随机向量中相应的第二随机分量，及该第二随机分量在相应的二进制向量中的位置信息，确定该第一分量对应的第三随机分量子值；基于该第一平方分量与该第三随机分量子值的和，确定所述第三距离中间结果；

所述第四距离中间结果是采用以下过程获得的：

针对所述第二向量中的每个第二分量，基于该第二分量对应的每个二进制分量及所述第一向量中相应的第一分量，确定该第二分量与相应的第一分

量的乘积；并基于所述第二随机向量中相应的第二随机分量、及该第二随机分量在相应的二进制向量中的位置信息，确定该第二分量对应的第四随机分量子值；基于该乘积与该第四随机分量子值的和，确定所述第四距离中间结果；

基于所述第二向量、所述第三距离中间结果和所述第四距离中间结果，确定所述第一向量和第二向量的目标距离，包括：

针对所述第二向量中的每个第二分量，确定该第二分量的第二平方分量；基于该第二分量的第二平方分量、该第二分量与相应的第一分量的乘积、第三随机分量子值、相应的第一分量的第一平方分量以及第四随机分量子值，确定该第二分量对应的子距离；

基于每个第二分量对应的子距离，确定所述第一向量和第二向量的目标距离。

11、根据权利要求 1 所述的方法，其中，所述基于所述目标距离以及预设的第一距离阈值，确定所述第一数据以及所述第二数据是否匹配包括：

判断所述目标距离是否小于预设的第一距离阈值；

若是，则确定所述第一数据与所述第二数据匹配；

否则，确定所述第一数据与所述第二数据不匹配。

12、根据权利要求 11 所述的方法，所述确定所述第一数据与所述第二数据匹配之后，所述方法还包括：

判断所述目标距离是否等于预设的第二距离阈值，若是，则确定所述第一数据与所述第二数据相同。

13、一种数据匹配方法，应用于第二设备，所述方法包括：

将待匹配的第二数据输入到预先训练完成的向量转化模型中，获得所述第二数据对应的第二向量；并基于不经意传输 OT 协议，生成第一随机向量；

基于所述第二向量、所述第一随机向量以及第二设定算法生成的第二距离中间结果，并将所述第二距离中间结果发送给第一设备，使第一设备基于第一向量、第一距离中间结果和所述第二距离中间结果，确定所述第一向量

和第二向量的目标距离；并使所述第一设备基于所述目标距离以及预设的第一距离阈值，确定所述第一数据以及所述第二数据是否匹配；其中，所述第一向量为将第一数据输入到所述第一设备中的预先训练完成的向量转化模型中获得的；所述第一距离中间结果为基于第一中间向量以及第一设定算法获得的，所述第一中间向量为基于所述第一向量对应的二进制向量、所述第二向量及所述第一随机向量生成的。

14、根据权利要求 13 所述的方法，其中，所述将待匹配的第二数据输入到预先训练完成的向量转化模型中，获得所述第二数据对应的第二向量包括:

确定所述第二数据对应的第二目标数据类型;

根据所述第二目标数据类型以及预先保存的数据类型和预先训练完成的向量转化模型的对应关系，确定所述第二数据对应的预先训练完成的第二目标向量转化模型;

将所述第二数据输入到所述预先训练完成的第二目标向量转化模型中，获得所述第二数据对应的所述第二向量。

15、根据权利要求 14 所述的方法，其中，所述第二目标数据类型为文本类型、数字类型、图像类型中的任意一种。

16、根据权利要求 15 所述的方法，其中，若所述第二目标数据类型为文本类型，所述第二目标数据类型对应的预先训练完成的第二目标向量转化模型为词向量模型或句向量模型；若所述第二目标数据类型为数字类型，所述第二目标数据类型对应的预先训练完成的第二目标向量转化模型为独热编码模型；若所述第二目标数据类型为图像类型，所述第二目标数据类型对应的预先训练完成的第二目标向量转化模型为图像向量模型。

17、根据权利要求 13 所述的方法，所述方法还包括:

接收所述第一设备发送的第三距离中间结果，所述第三距离中间结果为基于第一向量、第二随机向量以及所述第二设定算法生成的，其中，所述第二随机向量为所述第一设备基于所述 OT 协议生成的；并通过 OT 协议获取第二中间向量，其中所述第二中间向量是基于所述第二向量对应的二进制向量、

第一向量及所述第二随机向量生成的;

基于所述第二中间向量以及所述第一设定算法获取第四距离中间结果;

基于所述第二向量、所述第三距离中间结果和第四距离中间结果，确定所述第一向量和第二向量的目标距离;

基于所述目标距离以及预设的第一距离阈值，确定所述第一数据以及所述第二数据是否匹配。

18、根据权利要求 13 所述的方法，所述将所述第二距离中间结果发送给所述第一设备之后，所述方法还包括:

接收所述第一设备发送的目标距离，基于所述目标距离以及预设的第一距离阈值，确定所述第一数据以及所述第二数据是否匹配。

19、根据权利要求 13 所述的方法，所述方法还包括:

接收所述第一设备发送的所述第一数据以及所述第二数据是否匹配的匹配结果。

20、一种数据匹配装置，应用于第一设备，所述装置包括:

第一获取模块，用于将待匹配的第一数据输入到预先训练完成的向量转化模型中，获得所述第一数据对应的第一向量，并获得所述第一向量对应的二进制向量;

第一处理模块，用于通过不经意传输 OT 协议，获取第一中间向量;基于所述第一中间向量以及第一设定算法，获得第一距离中间结果;其中，所述第一中间向量是基于所述第一向量对应的二进制向量、第二向量及第一随机向量生成的，所述第二向量为第二设备将第二数据输入到所述第二设备中的预先训练完成的向量转化模型中获得的，所述第一随机向量为所述第二设备基于所述 OT 协议生成的;

所述第一处理模块，还用于接收所述第二设备发送的基于所述第二向量、所述第一随机向量以及第二设定算法生成的第二距离中间结果;基于所述第一向量、所述第一距离中间结果和所述第二距离中间结果，确定所述第一向量和第二向量的目标距离;

第一确定模块，用于基于所述目标距离以及预设的第一距离阈值，确定所述第一数据以及所述第二数据是否匹配。

21、一种数据匹配装置，应用于第二设备，所述装置包括：

第二获取模块，用于将待匹配的第二数据输入到预先训练完成的向量转化模型中，获得所述第二数据对应的第二向量；并基于不经意传输 OT 协议，生成第一随机向量；

第二处理模块，用于基于所述第二向量、所述第一随机向量以及第二设定算法生成的第二距离中间结果，并将所述第二距离中间结果发送给第一设备，使第一设备基于第一向量、第一距离中间结果和所述第二距离中间结果，确定所述第一向量和第二向量的目标距离；并使所述第一设备基于所述目标距离以及预设的第一距离阈值，确定所述第一数据以及所述第二数据是否匹配；其中，所述第一向量为将第一数据输入到所述第一设备中的预先训练完成的向量转化模型中获得的；所述第一距离中间结果为基于第一中间向量以及第一设定算法获得的，所述第一中间向量为基于所述第一向量对应的二进制向量、所述第二向量及所述第一随机向量生成的。

22、一种数据匹配系统，所述系统包括：

第一设备，用于将待匹配的第一数据输入到预先训练完成的向量转化模型中，获得所述第一数据对应的第一向量，并获得所述第一向量对应的二进制向量；

第二设备，用于将待匹配的第二数据输入到预先训练完成的向量转化模型中，获得所述第二数据对应的第二向量；将待匹配的第二数据输入到预先训练完成的向量转化模型中，获得所述第二数据对应的第二向量；并基于不经意传输 OT 协议，生成第一随机向量；

所述第一设备，还用于通过不经意传输 OT 协议，获取第一中间向量；基于所述第一中间向量以及第一设定算法，获得第一距离中间结果；其中，所述第一中间向量是基于所述第一向量对应的二进制向量、第二向量及第一随机向量生成的；

所述第二设备，还用于基于所述第二向量、所述第一随机向量以及第二设定算法生成的第二距离中间结果，并将所述第二距离中间结果发送给所述第一设备;

所述第一设备，还用于基于所述第一向量、所述第一距离中间结果和所述第二距离中间结果，确定所述第一向量和第二向量的目标距离；基于所述目标距离以及预设的第一距离阈值，确定所述第一数据以及所述第二数据是否匹配。

23、一种电子设备，所述电子设备包括处理器和存储器，所述存储器用于存储程序指令，所述处理器用于执行存储器中存储的计算机程序时实现权利要求 1-19 任一所述数据匹配方法的步骤。

24、一种计算机可读存储介质，其存储有计算机程序，所述计算机程序被处理器执行时实现权利要求 1-19 任一所述数据匹配方法的步骤。