一种推荐方法及相关装置

本申请要求于 2021 年 9 月 29 日提交中国专利局、申请号为 202111152705.1、发明名称为“一种推荐方法及相关装置”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

本申请涉及人工智能领域，尤其涉及一种推荐方法及相关装置。

背景技术

人工智能(artificial intelligence，AI)是利用数字计算机或者数字计算机控制的机器模拟、延伸和扩展人的智能，感知环境、获取知识并使用知识获得最佳结果的理论、方法、技术及应用系统。换句话说，人工智能是计算机科学的一个分支，它企图了解智能的实质，并生产出一种新的能以人类智能相似的方式作出反应的智能机器。人工智能也就是研究各种智能机器的设计原理与实现方法，使机器具有感知、推理与决策的功能。

选择率预测，是指预测用户在特定环境下对某个物品的选择概率。例如，应用商店、在线广告等应用的推荐系统中，选择率预测起到关键作用；通过选择率预测可以实现最大化企业的收益和提升用户满意度，推荐系统需同时考虑用户对物品的选择率和物品竞价，其中，选择率为推荐系统根据用户历史行为预测得到，而物品竞价代表该物品被选择/下载后系统的收益。例如，可以通过构建一个函数，该函数可以根据预测的用户选择率和物品竞价计算得到一个函数值，推荐系统按照该函数值对物品进行降序排列。

为了提高推荐系统的个性化程度，预测出符合用户偏好的物品列表，推荐系统通常将用户特征、商品特征、上下文特征等不同视角的特征进行交互，以此来捕获用户偏好。业界常用的特征交互方式主要分为两类。一类是堆叠式（stacked structure），一类是平行式（parallel structured）。

其中，平行式可以包括交叉网络（cross network）和深度网络（deep network），其中，交叉网络可以称之为显式交互网络（explicit component），深度网络可以称之为隐式交互网络（implicit component），在现有的实现中，交叉网络（cross network）和深度网络（deep network）将底层（embedding layer）输出的特征向量共同作为输入，交叉网络和深度网络各自独立处理数据（也就是只进行自身的数据交互过程），且相互之间不交互，两个网络在最后的输出层（output layer）进行融合并输出。代表的模型有 Wide & Deep，DCN，xDeepFM 等。

在当前平行式的交互模型中，由于平行的网络在各自的交互过程中没有任何信息共享，直到最后的输出层，才进行融合，可以称之为延迟融合（late fusion），这种方式忽略了不同特征交互方式之间的协同作用。且平行的网络之间输入的数据相同，忽略了特征针对不同交互方式的异质性，即不同的特征对不同的交互方式带来的信息量不同，导致了模型的数据处理精度较差。

发明内容

第一方面，本申请提供了一种推荐方法，所述方法包括：

获取目标特征向量，所述目标特征向量为对目标用户和目标物品的属性信息或进行特征提取得到的；

其中，目标用户的属性信息可以为与用户喜好特征相关的属性，性别、年龄、职业、收入、爱好以及受教育程度中的至少一种，其中，性别可以为男或者女，年龄可以为0-100之间的数字，职业可以为教师、程序员、厨师等等，爱好可以为篮球、网球、跑步等等，受教育程度可以为小学、初中、高中、大学等等；本申请并不限定目标用户的属性信息的具体类型。

其中，物品可以为实体物品，或者是虚拟物品，例如可以为APP、音视频、网页以及新闻资讯等物品，物品的属性信息可以为物品名称、开发者、安装包大小、品类以及好评度中的至少一种，其中，以物品为应用程序为例，物品的品类可以为聊天类、跑酷类游戏、办公类等等，好评度可以为针对于物品的打分、评论等；本申请并不限定物品的属性信息的具体类型。

通过推荐模型处理所述目标特征向量，以得到推荐信息，所述推荐信息用于表示所述目标用户对所述目标物品进行选择的概率；其中，所述推荐模型包括交叉网络（cross network）、深度网络（deep network）以及目标网络，所述交叉网络包括第一交叉层（cross layer）和第二交叉层，所述深度网络包括第一深度层（deep layer）和第二深度层；所述目标网络用于对所述第一交叉层输出的第一中间输出和所述第一深度层输出的第二中间输出进行融合处理，以得到第一融合结果，所述目标网络还用于处理所述第一融合结果，以得到所述第一交叉层对应的第一权重以及所述第一深度层对应的第二权重，并将所述第一融合结果分别与所述第一权重和所述第二权重进行加权处理，以得到第一中间输入和第二中间输入；所述第一中间输入为所述第二交叉层的输入数据，所述第二中间输入为所述第二深度层的输入数据；

其中，本申请实施例中的目标网络可以实现交叉网络和深度网络的网络层之间的数据交互，具体的，交叉网络可以包括多个交叉层，深度网络可以包括多个深度层，可选的，交叉网络中交叉层的数量和深度网络中深度层的数量一致（或者不一致，而是存在位置上的对应关系）。

例如，交叉网络中交叉层的数量和深度网络中深度层的数量一致时，则处于相同位置的交叉层和深度层是一一对应的。示例性的，交叉网络可以包括交叉层1、交叉层2、交叉层3、交叉层4、交叉层5，深度网络可以包括深度层1、深度层2、深度层3、深度层4、深度层5，则交叉层1对应于深度层1，交叉层2对应于深度层2，交叉层3对应于深度层3，交叉层4对应于深度层4，交叉层5对应于深度层5。

例如，交叉网络中交叉层的数量和深度网络中深度层的数量不一致时，则处于各自网络相对位置相同的交叉层和深度层是一一对应的。示例性的，交叉网络可以包括交叉层1、交叉层2、交叉层3，深度网络可以包括深度层1、深度层2、深度层3、深度层4、深度层5、深度层6，则交叉层1对应于深度层1和深度层2，交叉层2对应于深度层3和深度层4，交叉层

3对应于深度层5和深度层6。

在一种可能的实现中，所述交叉网络可以包括第一交叉层（cross layer）和第二交叉层，所述深度网络包括第一深度层（deep layer）和第二深度层；其中，第一交叉层和第二交叉层可以为交叉网络中任意相邻的网络层，第一深度层和第二深度层可以为深度网络中任意相邻的网络层。第一交叉层可以对应于第一深度层，第二交叉层可以对应于第二深度层，例如，交叉网络中交叉层的数量和深度网络中深度层的数量一致时，第一交叉层可以为交叉层1，第一深度层可以为深度层1，第二交叉层可以为交叉层2，第二深度层可以为深度层2。例如，交叉网络中交叉层的数量和深度网络中深度层的数量不一致时，第一交叉层可以为交叉层1，第一深度层可以为深度层1和深度层2，第二交叉层可以为交叉层2，第二深度层可以为深度层3和深度层4。

其中，所述目标网络用于对所述第一交叉层输出的第一中间输出和所述第一深度层输出的第二中间输出进行融合处理，以得到第一融合结果。

当所述推荐信息满足预设条件，确定向所述目标用户推荐所述目标物品。

本申请实施例中，目标网络可以将交叉网络中的交叉层的输出以及深度网络中的深度层的输出进行融合以及基于权重的适配，实现了交叉网络和深度网络之间的数据交互，提高了推荐模型的数据处理精度。

在一种可能的实现中，所述融合处理包括按点逐位相加（point-wise addition）、哈达玛积、拼接、基于注意力机制的池化中的一个。

其中，第一交叉层的输入可以为多个嵌入向量（或者称之为特征向量），第一交叉层输出的第一中间输出可以为多个嵌入向量，第一深度层的输入可以为多个嵌入向量，第一深度层输出的第二中间输出可以为多个嵌入向量，因此可以对第一交叉层输出的多个嵌入向量和第一深度层输出的多个嵌入向量进行融合处理，例如按点逐位相加、哈达玛积、拼接、基于注意力机制的池化。在一种可能的实现中，所述第一中间输出包括M个第一特征向量，所述第二中间输出包括M个第二特征向量，所述第一融合结果包括M个第三特征向量。

其中，所述目标网络还用于处理所述第一融合结果，以得到所述第一交叉层对应的第一权重以及所述第一深度层对应的第二权重，并将所述第一融合结果分别与所述第一权重和所述第二权重进行加权处理，以得到所述第一交叉层对应的第一中间输入和所述第一深度层对应的第二中间输入。

在一种可能的实现中，所述目标网络可以包括第一特征适配网络；所述第一特征适配网络为全连接网络、压缩奖惩网络（squeeze-and-excitation network）、注意力网络、SENet或门网络（gatenet），关于第一特征适配网络的描述可以参照上述实施例中关于第三特征适配网络的描述，这里不再赘述。进而目标网络可以通过所述第一特征适配网络处理所述第一融合结果。

在一种可能的实现中，所述第一中间输出包括M个第一特征向量，所述第二中间输出包括M个第二特征向量，所述第一融合结果包括M个第三特征向量，所述第一权重包括每

个第一特征向量对应的第一权重值，所述第二权重包括每个第二特征向量对应的第二权重值；在得到第一权重和第二权重之后，可以将所述M个第三特征向量中的每个第一特征向量与对应的第一权重值进行加权处理，将所述M个第三特征向量中的每个第二特征向量与对应的第二权重值进行加权处理。

其中，第一融合结果可以包括M个第三特征向量，第一权重可以包括各个第三特征向量对应的第一权重值，可选的，各个第三特征向量对应的第一权重值可以相同也可以不同，类似的，第二权重可以包括各个第三特征向量对应的第二权重值，可选的，各个第三特征向量对应的第二权重值可以相同也可以不同。

在一种可能的实现中，其特征在于，

所述交叉网络还包括第三交叉层，所述深度网络还包括第三深度层；所述目标网络还用于对所述第二交叉层输出的第三中间输出和所述第二深度层输出的第四中间输出进行融合处理，以得到第二融合结果，所述目标网络还用于处理所述第二融合结果，以得到所述第二交叉层对应的第三权重以及所述第二深度层对应的第四权重，并将所述第二融合结果分别与所述第三权重和所述第四权重进行加权处理，以得到所述第二交叉层对应的第三中间输入和所述第二深度层对应的第四中间输入；所述第三交叉层用于处理所述第三中间输入，所述第三深度层用于处理所述第四中间输入。

在一种可能的实现中，所述目标网络包括第二特征适配网络；所述第二特征适配网络为全连接网络、压缩奖惩网络（squeeze-and-excitation network）、注意力网络、SENet或门网络（gatenet）；

所述处理所述第二融合结果，包括：

通过所述第二特征适配网络处理所述第二融合结果。

在一种可能的实现中，在对各个嵌入向量进行拼接操作之后，可以不将拼接结果作为交叉网络和深度网络的输入，而是基于一个训练好的网络，该网络可以学习特征在交叉网络和深度网络上的权重分布，并基于权重分布调整拼接结果，以得到交叉网络和深度网络各自的输入。

在一种可能的实现中，参照图6a，上述网络可以为第三特征适配网络，可选的，第三特征适配网络可以包括两个子网络，一个子网络对应交叉网络，一个子网络对应深度网络，相当于为每一个特征交互网络（即交叉网络和深度网络）配置一个特征适配模块，学习特征在各个交互网络上的权重分布。

在一种可能的实现中，第三特征适配网络可以为全连接网络、压缩奖惩网络（squeeze-and-excitation network）、注意力网络、SENet或门网络（gatenet），其中，第三特征适配网络可以包括两个子网络，一个子网络对应交叉网络，一个子网络对应深度网络，子网络可以为全连接网络、压缩奖惩网络（squeeze-and-excitation network）、注意力网络、SENet或门网络（gatenet）。

具体的，可以基于嵌入层对目标用户和目标用户的各个属性信息分别进行特征提取，以得到每个属性信息对应的嵌入向量，可以对各个嵌入向量进行拼接操作（concat），以得到初始特征向量，并通过第三特征适配网络处理所述初始特征向量，以得到所述交叉网络对应的第五权重以及所述深度网络对应的第六权重，交叉网络对应的子网络可以处理初始特征向量，以得到所述交叉网络对应的第五权重，深度网络对应的子网络可以处理初始特征向量，以得到所述深度网络对应的第六权重。

其中，初始特征向量可以包括多个嵌入向量，第五权重可以包括各个嵌入向量对应的权重值，可选的，各个嵌入向量对应的权重值可以相同也可以不同，类似的，第六权重可以包括各个嵌入向量对应的权重值，可选的，各个嵌入向量对应的权重值可以相同也可以不同。

在一种可能的实现中，可以将所述初始特征向量分别与所述第五权重和所述第六权重进行加权处理，以得到所述交叉网络对应的第一网络输入和所述深度网络对应的第二网络输入，所述第一网络输入用于作为所述交叉网络的输入，所述第二网络输入用于作为所述深度网络的输入。相当于在输入层和特征交互层（即交叉网络和深度网络）之间，引入了第三特征适配网络，第三特征适配网络可以包括两个子网络，一个子网络对应交叉网络，一个子网络对应深度网络，相当于为每一个特征交互网络（即交叉网络和深度网络）配置一个特征适配模块，学习特征在各个交互网络上的权重分布，然后校准过后的特征（即第一网络输入和第二网络输入），分别输入到交叉网络和深度网络。

本申请实施例中，第三特征适配网络可以为不同的交互网络学习异质性的参数分布，避免过度的共享，进而将第一网络输入作为交叉网络的输入，将第二网络输入作为深度网络的输入，也可以提高推荐模型的数据处理精度。

在一种可能的实现中，所述用户属性包括如下的至少一种：性别，年龄，职业，收入，爱好，教育程度。

在一种可能的实现中，所述物品属性包括如下的至少一种：物品名称，开发者，安装包大小，品类，好评度。

第二方面，本申请提供了一种模型训练方法，所述方法包括：

获取目标特征向量，所述目标特征向量为对目标用户和目标物品的属性信息或进行特征提取得到的；

通过第一推荐模型处理所述目标特征向量，以得到推荐信息，所述推荐信息用于表示所述目标用户对所述目标物品进行选择的概率；其中，所述推荐模型包括交叉网络（cross network）、深度网络（deep network）以及目标网络，所述交叉网络包括第一交叉层（cross layer）和第二交叉层，所述深度网络包括第一深度层（deep layer）和第二深度层；所述目标网络用于对所述第一交叉层输出的第一中间输出和所述第一深度层输出的第二中间输出进行融合处理，以得到第一融合结果，所述目标网络还用于处理所述第一融合结果，以得

到所述第一交叉层对应的第一权重以及所述第一深度层对应的第二权重，并将所述第一融合结果分别与所述第一权重和所述第二权重进行加权处理，以得到第一中间输入和第二中间输入；所述第一中间输入为所述第二交叉层的输入数据，所述第二中间输入为所述第二深度层的输入数据；

根据所述目标用户对所述目标物品的真实选择结果、以及所述推荐信息，确定损失，并根据所述损失更新所述第一推荐模型，以得到第二推荐模型。

在一种可能的实现中，所述融合处理包括按点逐位相加（point-wise addition）、哈达玛积、拼接、基于注意力机制的池化中的一个。

在一种可能的实现中，所述目标网络包括第一特征适配网络；所述第一特征适配网络为全连接网络、压缩奖惩网络（squeeze-and-excitation network）、注意力网络、SENet或门网络（gatenet）；

所述处理所述第一融合结果，包括：

通过所述第一特征适配网络处理所述第一融合结果。

在一种可能的实现中，所述第一中间输出包括M个第一特征向量，所述第二中间输出包括M个第二特征向量，所述第一融合结果包括M个第三特征向量，所述第一权重包括每个第一特征向量对应的第一权重值，所述第二权重包括每个第二特征向量对应的第二权重值；

所述将所述第一融合结果分别与所述第一权重和所述第二权重进行加权处理，包括：

将所述M个第三特征向量中的每个第一特征向量与对应的第一权重值进行加权处理；

将所述M个第三特征向量中的每个第二特征向量与对应的第二权重值进行加权处理。

在一种可能的实现中，所述交叉网络还包括第三交叉层，所述深度网络还包括第三深度层；所述目标网络还用于对所述第二交叉层输出的第三中间输出和所述第二深度层输出的第四中间输出进行融合处理，以得到第二融合结果，所述目标网络还用于处理所述第二融合结果，以得到所述第二交叉层对应的第三权重以及所述第二深度层对应的第四权重，并将所述第二融合结果分别与所述第三权重和所述第四权重进行加权处理，以得到所述第二交叉层对应的第三中间输入和所述第二深度层对应的第四中间输入；所述第三交叉层用于处理所述第三中间输入，所述第三深度层用于处理所述第四中间输入。

在一种可能的实现中，所述目标网络包括第二特征适配网络；所述第二特征适配网络为全连接网络、压缩奖惩网络（squeeze-and-excitation network）、注意力网络、SENet或门网络（gatenet）；

所述处理所述第二融合结果，包括：

通过所述第二特征适配网络处理所述第二融合结果。

在一种可能的实现中，所述目标特征向量包括第一网络输入和第二网络输入；

所述获取目标特征向量，包括：

获取初始特征向量，所述初始特征向量为对目标用户和目标物品的属性信息或进行特征提取得到的；

通过第三特征适配网络处理所述初始特征向量，以得到所述交叉网络对应的第五权重以及所述深度网络对应的第六权重，并将所述初始特征向量分别与所述第五权重和所述第六权重进行加权处理，以得到所述交叉网络对应的第一网络输入和所述深度网络对应的第二网络输入，所述第一网络输入用于作为所述交叉网络的输入，所述第二网络输入用于作为所述深度网络的输入；

所述根据所述损失更新所述第一推荐模型，以得到第二推荐模型，包括：

根据所述损失更新所述第一推荐模型和所述第二特征适配网络，以得到第二推荐模型和更新后的所述第二特征适配网络。

在一种可能的实现中，所述第三特征适配网络为全连接网络、压缩奖惩网络、注意力网络、SENet或门网络。

在一种可能的实现中，所述用户属性包括如下的至少一种：性别，年龄，职业，收入，爱好，教育程度。

在一种可能的实现中，所述物品属性包括如下的至少一种：物品名称，开发者，安装包大小，品类，好评度。

第三方面，本申请提供了一种推荐装置，所述装置包括：

获取模块，用于获取目标特征向量，所述目标特征向量为对目标用户和目标物品的属性信息或进行特征提取得到的；

数据处理模块，用于通过推荐模型处理所述目标特征向量，以得到推荐信息，所述推荐信息用于表示所述目标用户对所述目标物品进行选择的概率；其中，所述推荐模型包括交叉网络（cross network）、深度网络（deep network）以及目标网络，所述交叉网络包括第一交叉层（cross layer）和第二交叉层，所述深度网络包括第一深度层（deep layer）和第二深度层；所述目标网络用于对所述第一交叉层输出的第一中间输出和所述第一深度层输出的第二中间输出进行融合处理，以得到第一融合结果，所述目标网络还用于处理所述第一融合结果，以得到所述第一交叉层对应的第一权重以及所述第一深度层对应的第二权重，并将所述第一融合结果分别与所述第一权重和所述第二权重进行加权处理，以得到第一中间输入和第二中间输入；所述第一中间输入为所述第二交叉层的输入数据，所述第二中间输入为所述第二深度层的输入数据；

推荐模块，用于当所述推荐信息满足预设条件，确定向所述目标用户推荐所述目标物

品。

在一种可能的实现中，所述融合处理包括按点逐位相加（point-wise addition）、哈达玛积、拼接、基于注意力机制的池化中的一个。

在一种可能的实现中，所述目标网络包括第一特征适配网络；所述第一特征适配网络为全连接网络、压缩奖惩网络（squeeze-and-excitation network）、注意力网络、SENet或门网络（gatenet）；

所述数据处理模块，具体用于：

通过所述第一特征适配网络处理所述第一融合结果。

在一种可能的实现中，所述第一中间输出包括M个第一特征向量，所述第二中间输出包括M个第二特征向量，所述第一融合结果包括M个第三特征向量，所述第一权重包括每个第一特征向量对应的第一权重值，所述第二权重包括每个第二特征向量对应的第二权重值；

所述数据处理模块，具体用于：

将所述M个第三特征向量中的每个第一特征向量与对应的第一权重值进行加权处理；

将所述M个第三特征向量中的每个第二特征向量与对应的第二权重值进行加权处理。

在一种可能的实现中，所述交叉网络还包括第三交叉层，所述深度网络还包括第三深度层；所述目标网络还用于对所述第二交叉层输出的第三中间输出和所述第二深度层输出的第四中间输出进行融合处理，以得到第二融合结果，所述目标网络还用于处理所述第二融合结果，以得到所述第二交叉层对应的第三权重以及所述第二深度层对应的第四权重，并将所述第二融合结果分别与所述第三权重和所述第四权重进行加权处理，以得到所述第二交叉层对应的第三中间输入和所述第二深度层对应的第四中间输入；所述第三交叉层用于处理所述第三中间输入，所述第三深度层用于处理所述第四中间输入。

在一种可能的实现中，所述目标网络包括第二特征适配网络；所述第二特征适配网络为全连接网络、压缩奖惩网络（squeeze-and-excitation network）、注意力网络、SENet或门网络（gatenet）；

所述数据处理模块，具体用于：

通过所述第二特征适配网络处理所述第二融合结果。

在一种可能的实现中，所述目标特征向量包括第一网络输入和第二网络输入；

所述获取模块，具体用于：

获取初始特征向量，所述初始特征向量为对目标用户和目标物品的属性信息或进行特征提取得到的；

通过第三特征适配网络处理所述初始特征向量，以得到所述交叉网络对应的第五权重以及所述深度网络对应的第六权重，并将所述初始特征向量分别与所述第五权重和所述第六权重进行加权处理，以得到所述交叉网络对应的第一网络输入和所述深度网络对应的第二网络输入，所述第一网络输入用于作为所述交叉网络的输入，所述第二网络输入用于作为所述深度网络的输入。

在一种可能的实现中，所述第三特征适配网络为全连接网络、压缩奖惩网络、注意力网络、SENet或门网络。

在一种可能的实现中，所述用户属性包括如下的至少一种：性别，年龄，职业，收入，爱好，教育程度。

所述物品属性包括如下的至少一种：物品名称，开发者，安装包大小，品类，好评度。

第四方面，本申请提供了一种模型训练装置，所述装置包括：

获取模块，用于获取目标特征向量，所述目标特征向量为对目标用户和目标物品的属性信息或进行特征提取得到的；

数据处理模块，用于通过第一推荐模型处理所述目标特征向量，以得到推荐信息，所述推荐信息用于表示所述目标用户对所述目标物品进行选择的概率；其中，所述推荐模型包括交叉网络（cross network）、深度网络（deep network）以及目标网络，所述交叉网络包括第一交叉层（cross layer）和第二交叉层，所述深度网络包括第一深度层（deep layer）和第二深度层；所述目标网络用于对所述第一交叉层输出的第一中间输出和所述第一深度层输出的第二中间输出进行融合处理，以得到第一融合结果，所述目标网络还用于处理所述第一融合结果，以得到所述第一交叉层对应的第一权重以及所述第一深度层对应的第二权重，并将所述第一融合结果分别与所述第一权重和所述第二权重进行加权处理，以得到第一中间输入和第二中间输入；所述第一中间输入为所述第二交叉层的输入数据，所述第二中间输入为所述第二深度层的输入数据；

模型训练模块，用于根据所述目标用户对所述目标物品的真实选择结果、以及所述推荐信息，确定损失，并根据所述损失更新所述第一推荐模型，以得到第二推荐模型。

在一种可能的实现中，所述融合处理包括按点逐位相加（point-wise addition）、哈达玛积、拼接、基于注意力机制的池化中的一个。

在一种可能的实现中，所述目标网络包括第一特征适配网络；所述第一特征适配网络为全连接网络、压缩奖惩网络（squeeze-and-excitation network）、注意力网络、SENet或门网络（gatenet）；

所述数据处理模块，具体用于：

通过所述第一特征适配网络处理所述第一融合结果。

在一种可能的实现中，所述第一中间输出包括M个第一特征向量，所述第二中间输出包括M个第二特征向量，所述第一融合结果包括M个第三特征向量，所述第一权重包括每个第一特征向量对应的第一权重值，所述第二权重包括每个第二特征向量对应的第二权重值；

所述数据处理模块，具体用于：

将所述M个第三特征向量中的每个第一特征向量与对应的第一权重值进行加权处理；

将所述M个第三特征向量中的每个第二特征向量与对应的第二权重值进行加权处理。

在一种可能的实现中，所述交叉网络还包括第三交叉层，所述深度网络还包括第三深度层；所述目标网络还用于对所述第二交叉层输出的第三中间输出和所述第二深度层输出的第四中间输出进行融合处理，以得到第二融合结果，所述目标网络还用于处理所述第二融合结果，以得到所述第二交叉层对应的第三权重以及所述第二深度层对应的第四权重，并将所述第二融合结果分别与所述第三权重和所述第四权重进行加权处理，以得到所述第二交叉层对应的第三中间输入和所述第二深度层对应的第四中间输入；所述第三交叉层用于处理所述第三中间输入，所述第三深度层用于处理所述第四中间输入。

在一种可能的实现中，所述目标网络包括第二特征适配网络；所述第二特征适配网络为全连接网络、压缩奖惩网络（squeeze-and-excitation network）、注意力网络、SENet或门网络（gatenet）；

所述数据处理模块，具体用于：

通过所述第二特征适配网络处理所述第二融合结果。

在一种可能的实现中，所述目标特征向量包括第一网络输入和第二网络输入；

所述获取模块，具体用于：

获取初始特征向量，所述初始特征向量为对目标用户和目标物品的属性信息或进行特征提取得到的；

通过第三特征适配网络处理所述初始特征向量，以得到所述交叉网络对应的第五权重以及所述深度网络对应的第六权重，并将所述初始特征向量分别与所述第五权重和所述第六权重进行加权处理，以得到所述交叉网络对应的第一网络输入和所述深度网络对应的第二网络输入，所述第一网络输入用于作为所述交叉网络的输入，所述第二网络输入用于作为所述深度网络的输入；

所述模型训练模块，具体用于：

根据所述损失更新所述第一推荐模型和所述第二特征适配网络，以得到第二推荐模型和更新后的所述第二特征适配网络。

在一种可能的实现中，所述第三特征适配网络为全连接网络、压缩奖惩网络、注意力网络、SENet或门网络。

在一种可能的实现中，所述用户属性包括如下的至少一种：性别，年龄，职业，收入，爱好，教育程度。

在一种可能的实现中，所述物品属性包括如下的至少一种：物品名称，开发者，安装包大小，品类，好评度。

第五方面，本申请实施例提供了一种推荐装置，可以包括存储器、处理器以及总线系统，其中，存储器用于存储程序，处理器用于执行存储器中的程序，以执行如上述第一方面任一可选的方法。

第六方面，本申请实施例提供了一种训练装置，可以包括存储器、处理器以及总线系统，其中，存储器用于存储程序，处理器用于执行存储器中的程序，以执行如上述第二方面任一可选的方法。

第七方面，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中存储有计算机程序，当其在计算机上运行时，使得计算机执行上述第一方面及任一可选的方法，上述第二方面及任一可选的方法。

第八方面，本申请实施例提供了一种计算机程序产品，包括代码，当代码被执行时，用于实现上述第一方面及任一可选的方法，上述第二方面及任一可选的方法。

第九方面，本申请提供了一种芯片系统，该芯片系统包括处理器，用于支持执行设备或训练设备实现上述方面中所涉及的功能，例如，发送或处理上述方法中所涉及的数据；或，信息。在一种可能的设计中，所述芯片系统还包括存储器，所述存储器，用于保存执行设备或训练设备必要的程序指令和数据。该芯片系统，可以由芯片构成，也可以包括芯片和其他分立器件。

本申请实施例提供了一种推荐方法，所述方法包括：获取目标特征向量，所述目标特征向量为对目标用户和目标物品的属性信息或进行特征提取得到的；通过推荐模型处理所述目标特征向量，以得到推荐信息，所述推荐信息用于表示所述目标用户对所述目标物品进行选择的概率；其中，所述推荐模型包括交叉网络（cross network）、深度网络（deep network）以及目标网络，所述交叉网络包括第一交叉层（cross layer）和第二交叉层，所述深度网络包括第一深度层（deep layer）和第二深度层；所述目标网络用于对所述第一交叉层输出的第一中间输出和所述第一深度层输出的第二中间输出进行融合处理，以得到第一融合结果，所述目标网络还用于处理所述第一融合结果，以得到所述第一交叉层对应的第一权重以及所述第一深度层对应的第二权重，并将所述第一融合结果分别与所述第一权重和所述第二权重进行加权处理，以得到第一中间输入和第二中间输入；所述第一中间输入为所述第二交叉层的输入数据，所述第二中间输入为所述第二深度层的输入数据；当所述推荐信息满足预设条件，确定向所述目标用户推荐所述目标物品。通过上述方式，目标网络为不同的

交互网络学习异质性的参数分布，避免过度的共享，引入不同交互网络之间的交互信号，增强多塔网络的协同作用，提高了模型的预测精度。目标网络可以将交叉网络中的交叉层的输出以及深度网络中的深度层的输出进行融合以及基于权重的适配，实现了交叉网络和深度网络之间的数据交互，提高了推荐模型的数据处理精度。

附图说明

图 1 为人工智能主体框架的一种结构示意图；
图 2 为本申请实施例提供的一种系统架构的示意图；
图 3 为本申请实施例提供的一种系统架构的示意图；
图 4 为本申请实施例提供的一种推荐流场景的示意图；
图 5 为本申请实施例提供的一种推荐方法的流程示意图；
图 6a 为一种推荐模型的示意；
图 6b 为一种推荐模型的示意；
图 7 为一种推荐模型的示意；
图 8 为一种推荐模型的示意；
图 9 为一种推荐模型的示意；
图 10 为本申请实施例提供的一种模型训练方法的流程示意图；
图 11 为本申请实施例提供的一种推荐装置的结构示意图；
图 12 为本申请实施例提供的一种模型训练装置的结构示意图；
图 13 为本申请实施例提供的一种执行设备的示意图；
图 14 为本申请实施例提供的一种训练设备的示意图；
图 15 为本申请实施例提供的一种芯片的示意图。

具体实施方式

下面结合本申请实施例中的附图对本申请实施例进行描述。本申请的实施方式部分使用的术语仅用于对本申请的具体实施例进行解释，而非旨在限定本申请。

下面结合附图，对本申请的实施例进行描述。本领域普通技术人员可知，随着技术的发展和新场景的出现，本申请实施例提供的技术方案对于类似的技术问题，同样适用。

本申请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术语在适当情况下可以互换，这仅仅是描述本申请的实施例中对相同属性的对象在描述时所采用的区分方式。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，以便包含一系列单元的过程、方法、系统、产品或设备不必限于那些单元，而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它单元。

首先对人工智能系统总体工作流程进行描述，请参见图 1，图 1 示出的为人工智能主体框架的一种结构示意图，下面从“智能信息链”（水平轴）和“IT 价值链”（垂直轴）两个维度对上述人工智能主题框架进行阐述。其中，“智能信息链”反映从数据的获取到处理的

一列过程。举例来说，可以是智能信息感知、智能信息表示与形成、智能推理、智能决策、智能执行与输出的一般过程。在这个过程中，数据经历了“数据—信息—知识—智慧”的凝练过程。“IT 价值链”从人智能的底层基础设施、信息（提供和处理技术实现）到系统的产业生态过程，反映人工智能为信息技术产业带来的价值。

（1）基础设施

基础设施为人工智能系统提供计算能力支持，实现与外部世界的沟通，并通过基础平台实现支撑。通过传感器与外部沟通；计算能力由智能芯片（CPU、NPU、GPU、ASIC、FPGA 等硬件加速芯片）提供；基础平台包括分布式计算框架及网络等相关的平台保障和支持，可以包括云存储和计算、互联互通网络等。举例来说，传感器和外部沟通获取数据，这些数据提供给基础平台提供的分布式计算系统中的智能芯片进行计算。

（2）数据

基础设施的上一层的数据用于表示人工智能领域的数据来源。数据涉及到图形、图像、语音、文本，还涉及到传统设备的物联网数据，包括已有系统的业务数据以及力、位移、液位、温度、湿度等感知数据。

（3）数据处理

数据处理通常包括数据训练，机器学习，深度学习，搜索，推理，决策等方式。

其中，机器学习和深度学习可以对数据进行符号化和形式化的智能信息建模、抽取、预处理、训练等。

推理是指在计算机或智能系统中，模拟人类的智能推理方式，依据推理控制策略，利用形式化的信息进行机器思维和求解问题的过程，典型的功能是搜索与匹配。

决策是指智能信息经过推理后进行决策的过程，通常提供分类、排序、预测等功能。

（4）通用能力

对数据经过上面提到的数据处理后，进一步基于数据处理的结果可以形成一些通用的能力，比如可以是算法或者一个通用系统，例如，翻译，文本的分析，计算机视觉的处理，语音识别，图像的识别等等。

（5）智能产品及行业应用

智能产品及行业应用指人工智能系统在各领域的产品和应用，是对人工智能整体解决方案的封装，将智能信息决策产品化、实现落地应用，其应用领域主要包括：智能终端、智能交通、智能医疗、自动驾驶、智慧城市等。

本申请实施例可以应用于信息推荐领域，该场景包括但不限于涉及电商产品推荐、搜索引擎结果推荐、应用市场推荐、音乐推荐、视频推荐等场景，各种不同应用场景中被推荐的物品也可以称为“对象”以方便后续描述，即在不同的推荐场景中，推荐对象可以是APP，或者视频，或者音乐，或者某款商品(如线上购物平台的呈现界面，会根据用户的不同而显示不同的商品进行呈现，这实质也可以是通过推荐模型的推荐结果来进行呈现)。这些推荐场景通常涉及用户行为日志采集、日志数据预处理(例如，量化、采样等)、样本集训练以获得推荐模型、根据推荐模型对训练样本项对应的场景中所涉及的对象(如 APP、音乐等)进行分析处理、例如，推荐模型训练环节中所选择的样本来自于手机应用市场用户对于

所推荐 APP 的操作行为，则由此所训练出来的推荐模型则适用于上述手机 APP 应用市场，或者可以用于其它的类型的终端的 APP 应用市场进行终端 APP 的推荐。推荐模型将最终计算出各个待推荐对象的推荐概率或者分值，推荐系统根据一定的选择规则选定的推荐结果，例如按照推荐概率或者分值进行排序，通过相应的应用或者终端设备呈现给用户、用户对推荐结果中的对象进行操作以生成用户行为日志等环节。

参照图 4，在推荐过程中，当一个用户与推荐系统进行交互会触发一个推荐请求，推荐系统会将该请求及其相关的特征信息输入到部署的推荐模型中，然后预测用户对所有候选对象的点击率。随后，根据预测的点击率对候选对象进行降序排列，按顺序将候选对象展示在不同的位置作为对用户的推荐结果。用户对展示的项目进行浏览并发生用户行为，如浏览、点击和下载等。这些用户行为会被存入日志中作为训练数据，通过离线训练模块不定期地更新推荐模型的参数，提高模型的推荐效果。

比如，用户打开手机应用市场即可触发应用市场的推荐模块，应用市场的推荐模块会根据用户的历史下载记录、用户点击记录，应用的自身特征，时间、地点等环境特征信息，预测用户对给定的各个候选应用的下载可能性。根据预测的结果，应用市场按照可能性降序展示，达到提高应用下载概率的效果。具体来说，将更有可能下载的应用排在靠前的位置，将不太可能下载的应用排列在靠后的位置。而用户的行为也会存入日志并通过离线训练模块对预测模型的参数进行训练和更新。

又比如，在终身伴侣相关的应用中，可以基于用户在视频、音乐、新闻等域的历史数据，通过各种模型和算法，仿照人脑机制，构建认知大脑，搭建用户终身学习系统框架。终身伴侣可以根据系统数据和应用数据等来记录用户过去发生的事件，理解用户的当前意图，预测用户未来的动作或行为，最终实现智能服务。在当前第一阶段，根据音乐APP、视频APP和浏览器APP等获取用户的行为数据（包含端侧短信、照片、邮件事件等信息），一方面构建用户画像系统，另一方面实现基于用户信息过滤、关联分析、跨域推荐、因果推理等的学习与记忆模块，构建用户个人知识图谱。

接下来介绍本申请实施例的应用架构。

参见附图2，本申请实施例提供了一种推荐系统架构200。数据采集设备260用于采集样本，一个训练样本可以由多个特征信息（或者描述为属性信息，例如用户属性以及物品属性）组成，特征信息可以有多种，具体可以包括用户特征信息和对象特征信息以及标签特征，用户特征信息用于表征用户的特征，例如性别，年龄，职业，爱好等，对象特征信息用于表征向用户所推送的对象的特征，不同的推荐系统对应不同的对象，不同的对象所需要提取的特征类型也不想同，例如APP市场的训练样本中所提取的对象特征可以为，APP的名称(标识)，类型，大小等；而电商类APP的训练样本中所提起的对象特征可以为，商品的名称，所属的类别，价格区间等；标签特征，则是用于表示这个样本是正例还是负例，通常样本的标签特征可以通过用户对所推荐对象的操作信息所获的，用户对所推荐对象有进行操作的样本为正例，用户对所推荐对象没有进行操作，或者仅浏览的样本为负例，例如当用户点击或者下载或者购买了所推荐的对象，则所述标签特征为1，表示该样本是正例，而如果用户没有对所推荐的对象进行任何操作，则所述标签特征为0，表示该样本是负例。

样本在采集后可以保存在数据库230中，数据库230中的样本中的部分或全部特征信息也可以直接从客户设备240中获取，如用户特征信息，用户对对象的操作信息(用于确定类型标识)，对象特征信息(如对象标识)等。训练设备220基于数据库230中样本训练获取模型参数矩阵用于生成推荐模型201。下面将更详细地描述训练设备220如何训练得到用于生成推荐模型201的模型参数矩阵，推荐模型201能够用于对大量对象进行评估从而得出各个待推荐对象的分值，进一步的还可以从大量对象的评估结果中推荐指定或者预设数目个对象，计算模块211基于推荐模型201的评估结果获取推荐结果，通过I/O接口 212推荐给客户设备。

在本申请实施例中，该训练设备 220 可以从数据库 230 中样本集内选取正、负样本添加到所述训练集中，之后采用推荐模型对训练集中的样本进行训练从而得到训练后的推荐模型；计算模块 211 的实现细节可以参照图 5 所示的方法实施例的详细描述。

训练设备 220 基于样本训练获得模型参数矩阵后用于构建推荐模型 201 后，将推荐模型 201 发送给执行设备 210，或者直接将模型参数矩阵发送给执行设备 210，在执行设备 210 中构建推荐模型，用于进行相应系统的推荐，例如基于视频相关的样本训练获得的推荐模型可以用于视频网站或 APP 中对用户进行视频的推荐，基于 APP 相关的样本训练获得的推荐模型可以用于应用市场中对用户进行 APP 的推荐。

执行设备 210 配置有 I/O 接口 212，与外部设备进行数据交互，执行设备 210 可以通过 I/O 接口 212 从客户设备 240 获取用户特征信息，例如用户标识、用户身份、性别、职业、爱好等，此部分信息也可以从系统数据库中获取。推荐模型 201 基于用户特征信息和待推荐对象特征信息向用户推荐目标推荐对象。执行设备 210 可以设置在云端服务器中，也可以设置于用户客户端中。

执行设备 210 可以调用数据存储系统 250 中的数据、代码等，同时也可以将输出的数据存入数据存储系统 250 中。数据存储系统 250 可以设置于执行设备 210 中，也可以独立设置，或者设置于其他网络实体中，数量可以是一个也可以是多个。

计算模块 211 使用推荐模型 201 对用户特征信息，待推荐对象特征信息进行处理，例如，该计算模块 211 使用推荐模型 201 对用户特征信息，以及待推荐对象的特征信息进行分析处理，从而得出该待推荐对象的分值，对待推荐对象按照分值进行排序，其中，排序靠前的对象将作为推荐给客户设备 240 的对象。

最后，I/O 接口 212 将推荐结果返回给客户设备 240，呈现给用户。

更深层地，训练设备 220 可以针对不同的目标，基于不同的样本特征信息生成相应的推荐模型 201，以给用户提供更佳的结果。

值得注意的，附图 2 仅是本申请实施例提供的一种系统架构的示意图，图中所示设备、器件、模块等之间的位置关系不构成任何限制，例如，在附图 2 中，数据存储系统 250 相对执行设备 210 是外部存储器，在其它情况下，也可将数据存储系统 250 置于执行设备 210 中。

在本申请实施例中，该训练设备 220、执行设备 210、客户设备 240 可以分别为三个不同的物理设备，也可能该训练设备 220 和执行设备 210 在同一个物理设备或者一个集群上，也可能该执行设备 210 与该客户设备 240 在同一个物理设备或者一个集群上。

参见附图 3，是本申请实施例提的一种系统架构 300。在此架构中执行设备 210 由一个或多个服务器实现，可选的，与其它计算设备配合，例如：数据存储、路由器、负载均衡器等设备；执行设备 210 可以布置在一个物理站点上，或者分布在多个物理站点上。执行设备 210 可以使用数据存储系统 250 中的数据，或者调用数据存储系统 250 中的程序代码实现对象推荐的功能，具体地，将待推荐的对象的信息输入到推荐模型中，推荐模型为每个待推荐对象生成预估分数，然后按照预估分数从高到低的顺序进行排序，按照排序结果向用户推荐该待推荐对象。例如，将排序结果中的前 10 个对象推荐给用户。

其中，数据存储系统 250 用于接收和存储训练设备发送的推荐模型的参数，以及用于存储通过推荐模型得到的推荐结果的数据，当然还可能包括该存储系统 250 正常运行所需的程序代码(或指令)。数据存储系统 250 可以为部署在执行设备 210 以外的一个设备或者多个设备构成的分布式存储集群，此时，当执行设备 210 需要使用存储系统 250 上的数据时，可以由存储系统 250 向执行设备 210 发送该执行设备所需的数据，相应地，该执行设备 210 接收并存储(或者缓存)该数据。当然数据存储系统 250 也可以部署在执行设备 210 内，当部署在执行设备 210 内时，该分布式存储系统可以包括一个或者多个存储器，可选的，存在多个存储器时，不同的存储器用于存储不同类型的数据，如通过训练设备生成的推荐模型的模型参数和通过推荐模型得到的推荐结果的数据可以分别存储在两个不同的存储器上。

用户可以操作各自的用户设备(例如本地设备 301 和本地设备 302)与执行设备 210 进行交互。每个本地设备可以表示任何计算设备，例如个人计算机、计算机工作站、智能手机、平板电脑、智能摄像头、智能汽车或其他类型蜂窝电话、媒体消费设备、可穿戴设备、机顶盒、游戏机等。

每个用户的本地设备可以通过任何通信机制/通信标准的通信网络与执行设备 210 进行交互，通信网络可以是广域网、局域网、点对点连接等方式，或它们的任意组合。

在另一种实现中，执行设备 210 可以由本地设备实现，例如，本地设备 301 可以基于推荐模型实现执行设备 210 的的推荐功能获取用户特征信息并向用户反馈推荐结果，或者为本地设备 302 的用户提供服务。

由于本申请实施例涉及大量神经网络的应用，为了便于理解，下面先对本申请实施例涉及的相关术语及神经网络等相关概念进行介绍。

1、点击概率(click-throughrate，CTR)

点击概率又可以称为点击率，是指网站或者应用程序上推荐信息(例如，推荐物品)被点击次数和曝光次数之比，点击率通常是推荐系统中衡量推荐系统的重要指标。

2、个性化推荐系统

个性化推荐系统是指根据用户的历史数据（例如本申请实施例中的操作信息），利用机器学习算法进行分析，并以此对新请求进行预测，给出个性化的推荐结果的系统。

3、离线训练(offlinetraining)

离线训练是指在个性化推荐系统中，根据用户的历史数据（例如本申请实施例中的操作信息），对推荐模型参数按照器学习的算法进行迭代更新直至达到设定要求的模块。

4、在线预测(onlineinference)

在线预测是指基于离线训练好的模型，根据用户、物品和上下文的特征预测该用户在当前上下文环境下对推荐物品的喜好程度，预测用户选择推荐物品的概率。

例如，图 3 是本申请实施例提供的推荐系统的示意图。如图 3 所示，当一个用户进入统，会触发一个推荐的请求，推荐系统会将该请求及其相关信息（例如本申请实施例中的操作信息）输入到推荐模型，然后预测用户对系统内的物品的选择率。进一步，根据预测的选择率或基于该选择率的某个函数将物品降序排列，即推荐系统可以按顺序将物品展示在不同的位置作为对用户的推荐结果。用户浏览不同的处于位置的物品并发生用户行为，如浏览、选择以及下载等。同时，用户的实际行为会存入日志中作为训练数据，通过离线训练模块不断更新推荐模型的参数，提高模型的预测效果。

例如，用户打开智能终端(例如，手机)中的应用市场即可触发应用市场中的推荐系统。应用市场的推荐系统会根据用户的历史行为日志，例如，用户的历史下载记录、用户选择记录，应用市场的自身特征，比如时间、地点等环境特征信息，预测用户下载推荐的各个候选 APP 的概率。根据计算的结果，应用市场的推荐系统可以按照预测的概率值大小降序展示候选 APP，从而提高候选 APP 的下载概率。

示例性地，可以将预测的用户选择率较高的 APP 展示在靠前的推荐位置，将预测的用户选择率较低的 APP 展示在靠后的推荐位置。

上述推荐模型可以是神经网络模型，下面对本申请实施例可能涉及的神经网络的相关术语和概念进行介绍。

(1)神经网络

神经网络可以是由神经单元组成的，神经单元可以是指以 xs（即输入数据）和截距 1 为输入的运算单元，该运算单元的输出可以为：

$$h_{W,b}(x)=f\left(W^{T}x\right)=f(\sum_{s=1}^{n}W_{s}x_{s}+b)$$

；

其中，s＝1、2、……n，n 为大于 1 的自然数，Ws 为 xs 的权重，b 为神经单元的偏置。f 为神经单元的激活函数(activation functions)，用于将非线性特性引入神经网络中，来将神经单元中的输入信号转换为输出信号。该激活函数的输出信号可以作为下一层卷积层的输入，激活函数可以是 sigmoid 函数。神经网络是将多个上述单一的神经单元联结在一起形成的网络，即一个神经单元的输出可以是另一个神经单元的输入。每个神经单元的输入可以与前一层的局部接受域相连，来提取局部接受域的特征，局部接受域可以是由若干个神经单元组成的区域。

（2）深度神经网络

深度神经网络（Deep Neural Network，DNN），也称多层神经网络，可以理解为具有很多层隐含层的神经网络，这里的“很多”并没有特别的度量标准。从 DNN 按不同层的位置划分，DNN 内部的神经网络可以分为三类：输入层，隐含层，输出层。一般来说第一层是输入层，最后一层是输出层，中间的层数都是隐含层。层与层之间是全连接的，也就是说，第 i 层的任意一个神经元一定与第 i+1 层的任意一个神经元相连。虽然 DNN 看起来很复杂，但是就每一层的工作来说，其实并不复杂，简单来说就是如下线性关系表达式：

$\vec{y} = \alpha(W\vec{x} + \vec{b})$，其中，$\vec{x}$是输入向量，$\vec{y}$是输出向量，$\vec{b}$是偏移向量，W是权重矩阵（也称系数），α（）是激活函数。每一层仅仅是对输入向量$\vec{x}$经过如此简单的操作得到输出向量$\vec{y}$。由于 DNN 层数多，则系数W和偏移向量$\vec{b}$的数量也就很多了。这些参数在 DNN 中的定义如下所述：以系数W为例：假设在一个三层的 DNN 中，第二层的第 4 个神经元到第三层的第 2 个神经元的线性系数定义为w_{24}^{3}。上标 3 代表系数W所在的层数，而下标对应的是输出的第三层索引 2 和输入的第二层索引 4。总结就是：第 L−1 层的第 k 个神经元到第 L 层的第 j 个神经元的系数定义为W_{jk}^{L}。需要注意的是，输入层是没有W参数的。在深度神经网络中，更多的隐含层让网络更能够刻画现实世界中的复杂情形。理论上而言，参数越多的模型复杂度越高，“容量”也就越大，也就意味着它能完成更复杂的学习任务。训练深度神经网络的也就是学习权重矩阵的过程，其最终目的是得到训练好的深度神经网络的所有层的权重矩阵（由很多层的向量 W 形成的权重矩阵）。

（3）损失函数

在训练深度神经网络的过程中，因为希望深度神经网络的输出尽可能的接近真正想要预测的值，所以可以通过比较当前网络的预测值和真正想要的目标值，再根据两者之间的差异情况来更新每一层神经网络的权重向量（当然，在第一次更新之前通常会有初始化的过程，即为深度神经网络中的各层预先配置参数），比如，如果网络的预测值高了，就调整权重向量让它预测低一些，不断的调整，直到深度神经网络能够预测出真正想要的目标值或与真正想要的目标值非常接近的值。因此，就需要预先定义“如何比较预测值和目标值之间的差异”，这便是损失函数（loss function）或目标函数（objective function），它们是用于衡量预测值和目标值的差异的重要方程。其中，以损失函数举例，损失函数的输出值（loss）越高表示差异越大，那么深度神经网络的训练就变成了尽可能缩小这个 loss 的过程。

（4）反向传播算法

可以采用误差反向传播（back propagation，BP）算法在训练过程中修正初始模型中参数的大小，使得模型的误差损失越来越小。具体地，前向传递输入信号直至输出会产生误差损失，通过反向传播误差损失信息来更新初始模型中的参数，从而使误差损失收敛。反向传播算法是以误差损失为主导的反向传播运动，旨在得到最优的模型参数，例如权重矩阵。

为了提高推荐系统的个性化程度，预测出符合用户偏好的物品列表，推荐系统通常将用户特征、商品特征、上下文特征等不同视角的特征进行交互，以此来捕获用户偏好。业界常用的特征交互方式主要分为两类。一类是堆叠式（stacked structure），一类是平行式（parallel structured）。

其中，平行式可以包括交叉网络（cross network）和深度网络（deep network），其中，交叉网络可以称之为显式交互网络（explicit component），深度网络可以称之为隐式交互网络（implicit component），在现有的实现中，交叉网络（cross network）和深度网络（deep network）将底层（embedding layer）输出的特征向量共同作为输入，交叉网络和深度网络

各自独立处理数据（也就是只进行自身的数据交互过程），且相互之间不交互，两个网络在最后的输出层（output layer）进行融合并输出。代表的模型有 Wide & Deep，DCN，xDeepFM 等。

在当前平行式的交互模型中，由于平行的网络在各自的交互过程中没有任何信息共享，直到最后的输出层，才进行融合，可以称之为延迟融合（late fusion），这种方式忽略了不同特征交互方式之间的协同作用。且平行的网络之间输入的数据相同，忽略了特征针对不同交互方式的异质性，即不同的特征对不同的交互方式带来的信息量不同，导致了模型的数据处理精度较差。

为了解决上述问题，本申请提供了一种推荐方法，接下来以模型推理阶段为例对本申请实施例提供的信息推荐方法进行说明。

参照图 5，图 5 为本申请实施例提供的一种推荐方法的实施例示意，如图 5 示出的那样，本申请实施例提供的一种推荐方法包括：

501、获取目标特征向量，所述目标特征向量为对目标用户和目标物品的属性信息或进行特征提取得到的。

本申请实施例中，步骤501的执行主体可以为终端设备，终端设备可以为便携式移动设备，例如但不限于移动或便携式计算设备(如智能手机)、个人计算机、服务器计算机、手持式设备(例如平板)或膝上型设备、多处理器系统、游戏控制台或控制器、基于微处理器的系统、机顶盒、可编程消费电子产品、移动电话、具有可穿戴或配件形状因子(例如，手表、眼镜、头戴式耳机或耳塞)的移动计算和/或通信设备、网络PC、小型计算机、大型计算机、包括上面的系统或设备中的任何一种的分布式计算环境等等。

本申请实施例中，步骤501的执行主体可以为云侧的服务器。

为了方便描述，以下不对执行主体的形态进行区分，都描述为执行设备。

在一种可能的实现中，为了计算目标用户对目标物品的选择概率，执行设备可以获取到目标用户的属性信息以及目标物品的属性信息。

其中，目标用户的属性信息可以为与用户喜好特征相关的属性，性别、年龄、职业、收入、爱好以及受教育程度中的至少一种，其中，性别可以为男或者女，年龄可以为0-100之间的数字，职业可以为教师、程序员、厨师等等，爱好可以为篮球、网球、跑步等等，受教育程度可以为小学、初中、高中、大学等等；本申请并不限定目标用户的属性信息的具体类型。

其中，物品可以为实体物品，或者是虚拟物品，例如可以为APP、音视频、网页以及新闻资讯等物品，物品的属性信息可以为物品名称、开发者、安装包大小、品类以及好评度中的至少一种，其中，以物品为应用程序为例，物品的品类可以为聊天类、跑酷类游戏、办公类等等，好评度可以为针对于物品的打分、评论等；本申请并不限定物品的属性信息的具体类型。

在一种可能的实现中，可以基于嵌入层对目标用户和目标物品的属性信息进行特征提取，以得到目标特征向量（特征向量也可以称之为嵌入向量）。

在一种可能的实现中，可以基于嵌入层（embedding layer）对目标用户和目标用户的各

个属性信息分别进行特征提取，以得到每个属性信息对应的嵌入向量，可以对各个嵌入向量进行拼接操作（concat），以得到目标特征向量，该目标特征向量可以作为交叉网络和深度网络的输入。

在一种可能的实现中，在对各个嵌入向量进行拼接操作之后，可以不将拼接结果作为交叉网络和深度网络的输入，而是基于一个训练好的网络，该网络可以学习特征在交叉网络和深度网络上的权重分布，并基于权重分布调整拼接结果，以得到交叉网络和深度网络各自的输入。

在一种可能的实现中，参照图6a，上述网络可以为第三特征适配网络，可选的，第三特征适配网络可以包括两个子网络，一个子网络对应交叉网络，一个子网络对应深度网络，相当于为每一个特征交互网络（即交叉网络和深度网络）配置一个特征适配模块，学习特征在各个交互网络上的权重分布。

在一种可能的实现中，第三特征适配网络可以为全连接网络、压缩奖惩网络（squeeze-and-excitation network）、注意力网络、SENet或门网络（gatenet），其中，第三特征适配网络可以包括两个子网络，一个子网络对应交叉网络，一个子网络对应深度网络，子网络可以为全连接网络、压缩奖惩网络（squeeze-and-excitation network）、注意力网络、SENet或门网络（gatenet）。

具体的，可以基于嵌入层对目标用户和目标用户的各个属性信息分别进行特征提取，以得到每个属性信息对应的嵌入向量，可以对各个嵌入向量进行拼接操作（concat），以得到初始特征向量，并通过第三特征适配网络处理所述初始特征向量，以得到所述交叉网络对应的第五权重以及所述深度网络对应的第六权重，交叉网络对应的子网络可以处理初始特征向量，以得到所述交叉网络对应的第五权重，深度网络对应的子网络可以处理初始特征向量，以得到所述深度网络对应的第六权重。

其中，初始特征向量可以包括多个嵌入向量，第五权重可以包括各个嵌入向量对应的权重值，可选的，各个嵌入向量对应的权重值可以相同也可以不同，类似的，第六权重可以包括各个嵌入向量对应的权重值，可选的，各个嵌入向量对应的权重值可以相同也可以不同。

在一种可能的实现中，可以将所述初始特征向量分别与所述第五权重和所述第六权重进行加权处理，以得到所述交叉网络对应的第一网络输入和所述深度网络对应的第二网络输入，所述第一网络输入用于作为所述交叉网络的输入，所述第二网络输入用于作为所述深度网络的输入。相当于在输入层和特征交互层（即交叉网络和深度网络）之间，引入了第三特征适配网络，第三特征适配网络可以包括两个子网络，一个子网络对应交叉网络，一个子网络对应深度网络，相当于为每一个特征交互网络（即交叉网络和深度网络）配置一个特征适配模块，学习特征在各个交互网络上的权重分布，然后校准过后的特征（即第一网络输入和第二网络输入），分别输入到交叉网络和深度网络。

本申请实施例中，第三特征适配网络可以为不同的交互网络学习异质性的参数分布，避免过度的共享，进而将第一网络输入作为交叉网络的输入，将第二网络输入作为深度网络的输入，也可以提高推荐模型的数据处理精度。

502、通过推荐模型处理所述目标特征向量，以得到推荐信息，所述推荐信息用于表示所述目标用户对所述目标物品进行选择的概率；其中，所述推荐模型包括交叉网络（cross network）、深度网络（deep network）以及目标网络，所述交叉网络包括第一交叉层（cross layer）和第二交叉层，所述深度网络包括第一深度层（deep layer）和第二深度层；所述目标网络用于对所述第一交叉层输出的第一中间输出和所述第一深度层输出的第二中间输出进行融合处理，以得到第一融合结果，所述目标网络还用于处理所述第一融合结果，以得到所述第一交叉层对应的第一权重以及所述第一深度层对应的第二权重，并将所述第一融合结果分别与所述第一权重和所述第二权重进行加权处理，以得到第一中间输入和第二中间输入；所述第一中间输入为所述第二交叉层的输入数据，所述第二中间输入为所述第二深度层的输入数据。

在一种可能的实现中，可以将目标特征向量作为推荐模型的输入，可选的，可以将目标特征向量作为推荐模型中交叉网络和深度网络的输入，例如可以将上述第一网络输入作为交叉网络的输入，将第二网络输入作为深度网络的输入。

接下来介绍交叉网络：

参照图6b，图6b为交叉网络的一个结构示意，其中交叉网络设计了一个固定的交互模式，即每次都将上一层的交互结果和输入层做内积，并在结果上加上上一层的交互结果。重复此种交互方式越多，交互的阶数也越多。交叉网络可以包括多个交叉层，参照图7，图7示出了交叉层的交互方式，其中x_0为输入层，x'初始为 x_0 ，之后为上一层的输出（也就是 y），w为权重参数，b为权重偏移量，x为上一层的输除（也就是y），可以理解成和x'一样。

接下来介绍深度网络：

由于交叉网络的参数数目少，限制了模型的能力（capacity）。为了捕获高阶非线性交叉，可以平行引入一个深度网络。可选的，深度网络可以为一个全连接的前馈神经网络。参照图8，图8为深度网络的一个结构示意。

在现有的实现中，交叉网络和深度网络之间不存在数据交互，本申请实施例中，可以在交叉网络和深度网络之间引入一个用于进行数据交互的目标网络。

接下里介绍本申请实施例中的目标网络：

其中，本申请实施例中的目标网络可以实现交叉网络和深度网络的网络层之间的数据交互，具体的，交叉网络可以包括多个交叉层，深度网络可以包括多个深度层，可选的，交叉网络中交叉层的数量和深度网络中深度层的数量一致（或者不一致，而是存在位置上的对应关系）。

例如，交叉网络中交叉层的数量和深度网络中深度层的数量一致时，则处于相同位置的交叉层和深度层是一一对应的。示例性的，交叉网络可以包括交叉层1、交叉层2、交叉层3、交叉层4、交叉层5，深度网络可以包括深度层1、深度层2、深度层3、深度层4、深度层5，则交叉层1对应于深度层1，交叉层2对应于深度层2，交叉层3对应于深度层3，交叉层4对应于深度层4，交叉层5对应于深度层5。

例如，交叉网络中交叉层的数量和深度网络中深度层的数量不一致时，则处于各自网络相对位置相同的交叉层和深度层是一一对应的。示例性的，交叉网络可以包括交叉层1、交叉层2、交叉层3，深度网络可以包括深度层1、深度层2、深度层3、深度层4、深度层5、深度层6，则交叉层1对应于深度层1和深度层2，交叉层2对应于深度层3和深度层4，交叉层3对应于深度层5和深度层6。

在一种可能的实现中，所述交叉网络可以包括第一交叉层（cross layer）和第二交叉层，所述深度网络包括第一深度层（deep layer）和第二深度层；其中，第一交叉层和第二交叉层可以为交叉网络中任意相邻的网络层，第一深度层和第二深度层可以为深度网络中任意相邻的网络层。第一交叉层可以对应于第一深度层，第二交叉层可以对应于第二深度层，例如，交叉网络中交叉层的数量和深度网络中深度层的数量一致时，第一交叉层可以为交叉层1，第一深度层可以为深度层1，第二交叉层可以为交叉层2，第二深度层可以为深度层2。例如，交叉网络中交叉层的数量和深度网络中深度层的数量不一致时，第一交叉层可以为交叉层1，第一深度层可以为深度层1和深度层2，第二交叉层可以为交叉层2，第二深度层可以为深度层3和深度层4。

其中，所述目标网络用于对所述第一交叉层输出的第一中间输出和所述第一深度层输出的第二中间输出进行融合处理，以得到第一融合结果。（目标网络处理数据的示意可以参照图9所示）。

在一种可能的实现中，所述融合处理可以包括按点逐位相加（point-wise addition）、哈达玛积、拼接、基于注意力机制的池化中的一个。

其中，第一交叉层的输入可以为多个嵌入向量（或者称之为特征向量），第一交叉层输出的第一中间输出可以为多个嵌入向量，第一深度层的输入可以为多个嵌入向量，第一深度层输出的第二中间输出可以为多个嵌入向量，因此可以对第一交叉层输出的多个嵌入向量和第一深度层输出的多个嵌入向量进行融合处理，例如按点逐位相加、哈达玛积、拼接、基于注意力机制的池化。在一种可能的实现中，所述第一中间输出包括M个第一特征向量，所述第二中间输出包括M个第二特征向量，所述第一融合结果包括M个第三特征向量。

其中，所述目标网络还用于处理所述第一融合结果，以得到所述第一交叉层对应的第一权重以及所述第一深度层对应的第二权重，并将所述第一融合结果分别与所述第一权重和所述第二权重进行加权处理，以得到所述第一交叉层对应的第一中间输入和所述第一深度层对应的第二中间输入。

在一种可能的实现中，所述目标网络可以包括第一特征适配网络；所述第一特征适配网络为全连接网络、压缩奖惩网络（squeeze-and-excitation network）、注意力网络、SENet或门网络（gatenet），关于第一特征适配网络的描述可以参照上述实施例中关于第三特征适配网络的描述，这里不再赘述。进而目标网络可以通过所述第一特征适配网络处理所述第一融合结果。

在一种可能的实现中，所述第一中间输出包括M个第一特征向量，所述第二中间输出包括M个第二特征向量，所述第一融合结果包括M个第三特征向量，所述第一权重包括每个第一特征向量对应的第一权重值，所述第二权重包括每个第二特征向量对应的第二权重

值；在得到第一权重和第二权重之后，可以将所述M个第三特征向量中的每个第一特征向量与对应的第一权重值进行加权处理，将所述M个第三特征向量中的每个第二特征向量与对应的第二权重值进行加权处理。

其中，第一融合结果可以包括M个第三特征向量，第一权重可以包括各个第三特征向量对应的第一权重值，可选的，各个第三特征向量对应的第一权重值可以相同也可以不同，类似的，第二权重可以包括各个第三特征向量对应的第二权重值，可选的，各个第三特征向量对应的第二权重值可以相同也可以不同。

在一种可能的实现中，可以将所述第一融合结果分别与所述第一权重和所述第二权重进行加权处理，以得到所述第一交叉层对应的第一中间输入和所述第一深度层对应的第二中间输入，所述第一中间输入用于作为所述第二交叉层的输入，所述第二中间输入用于作为所述第二深度层的输入。相当于在交叉网络和深度网络之间，引入了第一特征适配网络，第一特征适配网络可以包括两个子网络，一个子网络对应交叉网络，一个子网络对应深度网络，相当于为每一个特征交互网络（即交叉网络和深度网络）配置一个特征适配模块，学习特征在各个交互网络上的权重分布，然后校准过后的特征（即第一中间输入和第二中间输入），分别输入到下一个交叉层和深度层。例如，在得到第一中间输入和第二中间输入之后，所述第二交叉层可以处理所述第一中间输入，所述第二深度层可以处理所述第二中间输入。

本申请实施例中，目标网络可以将交叉网络中的交叉层的输出以及深度网络中的深度层的输出进行融合以及基于权重的适配，实现了交叉网络和深度网络之间的数据交互，提高了推荐模型的数据处理精度。

在一种可能的实现中，目标网络还可以对第二交叉层和第二深度层的输出进行融合以及基于权重的适配，例如，所述交叉网络还包括第三交叉层，所述深度网络还包括第三深度层；所述目标网络还用于对所述第二交叉层输出的第三中间输出和所述第二深度层输出的第四中间输出进行融合处理，以得到第二融合结果，所述目标网络还用于处理所述第二融合结果，以得到所述第二交叉层对应的第三权重以及所述第二深度层对应的第四权重，并将所述第二融合结果分别与所述第三权重和所述第四权重进行加权处理，以得到所述第二交叉层对应的第三中间输入和所述第二深度层对应的第四中间输入；所述第三交叉层用于处理所述第三中间输入，所述第三深度层用于处理所述第四中间输入。

在一种可能的实现中，所述目标网络包括第二特征适配网络；所述第二特征适配网络为全连接网络、压缩奖惩网络（squeeze-and-excitation network）、注意力网络、SENet或门网络（gatenet）；进而可以通过所述第二特征适配网络处理所述第二融合结果。

在一种可能的实现中，所述第二特征适配网络为全连接网络、压缩奖惩网络、注意力网络、SENet或门网络。

如此往复直至输出层。最后可以将不同交互网络的输出和融合后的输出进行融合，经过激活函数，最终得到推荐信息，所述推荐信息用于表示所述目标用户对所述目标物品进行选择的概率（即预测值y^）。

503、当所述推荐信息满足预设条件，确定向所述目标用户推荐所述目标物品。

通过上述方式，可以得到目标用户进行针对于目标物品的选择的概率，并基于上述概率进行信息推荐，具体的，当推荐信息满足预设条件，可以确定向所述目标用户推荐所述目标物品。

接下来描述预设条件：

在一种可能的实现中，在对目标用户进行信息推荐时，可以计算得到目标用户对多个物品（包括目标物品）进行选择的概率，并基于户对多个物品（包括目标物品）进行选择的概率来确定各个物品的对于该目标用户的推荐指数。

在得到各个物品的对于该目标用户的推荐指数之后，可以对推荐指数进行排序，并向目标用户推荐推荐指数最大的M个物品。

在一种可能的实现中，还可以选择可以设置一个概率阈值，当目标用户对多个物品（包括目标物品）进行选择的概率大于上述概率阈值，就可以向所述目标用户推荐。

在进行信息推荐时，可以以列表页的形式将推荐信息推荐给用户，以期望用户进行行为动作。

接下来以智能助手服务直达中的点击率预测场景为例介绍本申请实施例中的技术方案。点击率预测模型，其输入包括用户特征、商品特征和上下文特征，模型通过显式或隐式的方式对这些特征进行交互，增强多种交互方式之间的信息共享和协同作用是本专利技术的核心。离线训练点击率预测模型时，具体流程如下所示：为每一个特征交互网络学习不同的特征分布，并分别输入多种交互网络，将通过不同交互网络交互的结果进行融合，共享不同网络学习到的交互模式，根据融合后的网络参数，为后面不同的交互网络学习异质性的特征分布。循环上述步骤，直至输出层。将不同交互网络的输出、以及融合后的结果进行拼接，输入激活函数，最终得到预测值。在线服务时，直接加载模型进行线上预估。

接下来结合试验描述本申请实施例的有益效果，使用三个数据集：Criteo 数据集、Avazu 数据集和华为工业数据集，统计信息如表 1 所示。

Table 1: Statistics of evaluation datasets.

Dataset	#Feature Fields	#Instances ($\times 10^7$)	Positive Ratio
Avazu	23	3.64	17%
Criteo	39	9.69	50%
Industrial	85	8.75	3.3%

实验评价指标离线为 AUC，log loss；线上为点击率预测（click through rate，CTR），ECPM，在三个数据集上进行实验，以 DCN 作为模型基本骨架为例，实验结果如表 2 所示。从表中可以看出，相比于对比基线，本申请实施例可以取得最好的结果。

Model	Avazu		Criteo		Industrial	
	AUC	LogLoss	AUC	LogLoss	AUC	LogLoss
FNN	0.7738	0.3841	0.7941	0.5482	0.7261	0.1368
Wide&Deep	0.7745	0.3836	0.7952	0.5470	0.7255	0.1369
DeepFM	0.7747	0.3833	0.7955	0.5467	0.7262	0.1369
DCN	0.7751	0.3835	0.7963	0.5459	0.7263	0.1368
DCN-V2	0.7755	0.3822	0.7983	0.5435	0.7272	0.1368
xDeepFM	0.7754	0.3825	0.7968	0.5455	0.7275	0.1367
AutoInt	0.7756	0.3821	0.7983	0.5437	0.7282	0.1365
FNN	0.7759	0.3820	0.7985	0.5434	0.7269	0.1366
EDCN	**0.7793***	**0.3803***	**0.8001***	**0.5415***	**0.7310***	**0.1361***
Rel Impr.	0.44%	0.45%	0.20%	0.35%	0.38%	0.29%

本申请实施例可以是一个通用的特征交互增强框架，可以提升不同多塔模型的推荐效果，选取了几种工业常用的深度模型，用于进行 CTR 预测，并为这些模型引入本申请实施例涉及到的模块，以验证其通用性。实验结果如表 3 及表 4 所示，其中，多塔交互模块为本申请实施例中描述的目标网络。

表 3 特征适配模块的通用性

Model	Avazu		Criteo		Industrial	
	AUC	LogLoss	AUC	LogLoss	AUC	LogLoss
DeepFM	0.7747	0.3833	0.7955	0.5467	0.7262	0.1369
$\text{DeepFM}_{\text{Regulation}}$	0.7756	0.3831	0.7968	0.5452	0.7280	0.1365
xDeepFM	0.7754	0.3825	0.7968	0.5455	0.7275	0.1367
$\text{xDeepFM}_{\text{Regulation}}$	0.7765	0.3817	0.7977	0.5443	0.7288	0.1365
AutoInt	0.7756	0.3821	0.7983	0.5437	0.7282	0.1365
$\text{AutoInt}_{\text{Regulation}}$	0.7763	0.3818	0.7986	0.5432	0.7294	0.1365
DCN	0.7751	0.3835	0.7963	0.5459	0.7263	0.1368
$\text{DCN}_{\text{Regulation}}$	0.7771	0.3817	0.7985	0.5433	0.7286	0.1365
DCN-V2	0.7755	0.3822	0.7983	0.5435	0.7272	0.1368
$\text{DCN-V2}_{\text{Regulation}}$	0.7761	0.3819	0.7985	0.5433	0.7289	0.1365
$\text{DCN}_{\text{Bridge}}$	0.7775	0.3814	0.7997	0.5420	0.7300	0.1364
$\text{DCN}_{\text{BridgeRegulate}}$	0.7793	0.3803	0.8001	0.5415	0.7310	0.1361

Note that model $\text{DCN}_{\text{BridgeRegulate}}$ = EDCN.

表 4 多塔交互模块的通用性

Model	Avazu		Criteo		Industrial	
	AUC	LogLoss	AUC	LogLoss	AUC	LogLoss
xDeepFM	0.7754	0.3825	0.7968	0.5455	0.7275	0.1367
$\text{xDeepFM}_{\text{Bridge}}$	0.7761	0.3824	0.7974	0.5445	0.7291	0.1364
DCN	0.7751	0.3835	0.7963	0.5459	0.7263	0.1368
$\text{DCN}_{\text{Bridge}}$	0.7775	0.3814	0.7997	0.5420	0.7300	0.1364
DCN-V2	0.7755	0.3822	0.7983	0.5435	0.7272	0.1368
$\text{DCN-V2}_{\text{Bridge}}$	0.7776	0.3811	0.7987	0.5431	0.7292	0.1364

可以看出这些常用的多塔深度模型中引入本申请专利的模块后，可以显著提升模型性能。显示出本框架具有良好的兼容性。

本申请实施例提供了一种推荐方法，所述方法包括：获取目标特征向量，所述目标特征向量为对目标用户和目标物品的属性信息或进行特征提取得到的；通过推荐模型处理所述目标特征向量，以得到推荐信息，所述推荐信息用于表示所述目标用户对所述目标物品进行选择的概率；其中，所述推荐模型包括交叉网络（cross network）、深度网络（deep network）以及目标网络，所述交叉网络包括第一交叉层（cross layer）和第二交叉层，所述深度网络包括第一深度层（deep layer）和第二深度层；所述目标网络用于对所述第一交叉层输出的第一中间输出和所述第一深度层输出的第二中间输出进行融合处理，以得到第一融合结果，所述目标网络还用于处理所述第一融合结果，以得到所述第一交叉层对应的第一权重以及所述第一深度层对应的第二权重，并将所述第一融合结果分别与所述第一权重和所述第二权重进行加权处理，以得到第一中间输入和第二中间输入；所述第一中间输入为所述第二交叉层的输入数据，所述第二中间输入为所述第二深度层的输入数据；当所述推荐信息满足预设条件，确定向所述目标用户推荐所述目标物品。通过上述方式，目标网络为不同的交互网络学习异质性的参数分布，避免过度的共享，引入不同交互网络之间的交互信号，增强多塔网络的协同作用，提高了模型的预测精度。目标网络可以将交叉网络中的交叉层的输出以及深度网络中的深度层的输出进行融合以及基于权重的适配，实现了交叉网络和深度网络之间的数据交互，提高了推荐模型的数据处理精度。

以上从模型的推理过程对本申请实施例提供的推荐方法进行了描述，接下来从模型的训练过程进行描述。

参照图10，图10为本申请实施例提供的一种模型训练方法的流程示意，如图10所示，本申请实施例提供的一种模型训练方法包括：

1001、获取目标特征向量，所述目标特征向量为对目标用户和目标物品的属性信息或进行特征提取得到的。

1002、通过第一推荐模型处理所述目标特征向量，以得到推荐信息，所述推荐信息用于表示所述目标用户对所述目标物品进行选择的概率；其中，所述推荐模型包括交叉网络（cross network）、深度网络（deep network）以及目标网络，所述交叉网络包括第一交叉层（cross layer）和第二交叉层，所述深度网络包括第一深度层（deep layer）和第二深度层；所述目标网络用于对所述第一交叉层输出的第一中间输出和所述第一深度层输出的第二中间输出进行融合处理，以得到第一融合结果，所述目标网络还用于处理所述第一融合结果，以得到所述第一交叉层对应的第一权重以及所述第一深度层对应的第二权重，并将所述第一融合结果分别与所述第一权重和所述第二权重进行加权处理，以得到第一中间输入和第二中间输入；所述第一中间输入为所述第二交叉层的输入数据，所述第二中间输入为所述第二深度层的输入数据；

1003、根据所述目标用户对所述目标物品的真实选择结果、以及所述推荐信息，确定损失，并根据所述损失更新所述第一推荐模型，以得到第二推荐模型。

在一种可能的实现中，可以采用标签数据y和预测值y^，基于交叉检验熵（LogLoss）、平方差均值（RMSE）等损失函数得到loss，基于该loss和梯度下降算法，利用链式法则，

即可完成自动特征离散化模块、深度模型模型等不同模块参数的联合训练和优化。通过模型的损失函数不断调整多塔交互模块和特征适配模块的参数，最终得到优化后的模块。

在一种可能的实现中，所述融合处理包括按点逐位相加（point-wise addition）、哈达玛积、拼接、基于注意力机制的池化中的一个。

在一种可能的实现中，所述目标网络包括第一特征适配网络；所述第一特征适配网络为全连接网络、压缩奖惩网络（squeeze-and-excitation network）、注意力网络、SENet或门网络（gatenet）；可以通过所述第一特征适配网络处理所述第一融合结果。

在一种可能的实现中，所述第一中间输出包括M个第一特征向量，所述第二中间输出包括M个第二特征向量，所述第一融合结果包括M个第三特征向量，所述第一权重包括每个第一特征向量对应的第一权重值，所述第二权重包括每个第二特征向量对应的第二权重值；可以将所述M个第三特征向量中的每个第一特征向量与对应的第一权重值进行加权处理；将所述M个第三特征向量中的每个第二特征向量与对应的第二权重值进行加权处理。

在一种可能的实现中，所述交叉网络还包括第三交叉层，所述深度网络还包括第三深度层；所述目标网络还用于对所述第二交叉层输出的第三中间输出和所述第二深度层输出的第四中间输出进行融合处理，以得到第二融合结果，所述目标网络还用于处理所述第二融合结果，以得到所述第二交叉层对应的第三权重以及所述第二深度层对应的第四权重，并将所述第二融合结果分别与所述第三权重和所述第四权重进行加权处理，以得到所述第二交叉层对应的第三中间输入和所述第二深度层对应的第四中间输入；所述第三交叉层用于处理所述第三中间输入，所述第三深度层用于处理所述第四中间输入。

在一种可能的实现中，所述目标网络包括第二特征适配网络；所述第二特征适配网络为全连接网络、压缩奖惩网络（squeeze-and-excitation network）、注意力网络、SENet或门网络（gatenet）；可以通过所述第二特征适配网络处理所述第二融合结果。

在一种可能的实现中，所述目标特征向量包括第一网络输入和第二网络输入；可以获取初始特征向量，所述初始特征向量为对目标用户和目标物品的属性信息或进行特征提取得到的；通过第三特征适配网络处理所述初始特征向量，以得到所述交叉网络对应的第五权重以及所述深度网络对应的第六权重，并将所述初始特征向量分别与所述第五权重和所述第六权重进行加权处理，以得到所述交叉网络对应的第一网络输入和所述深度网络对应的第二网络输入，所述第一网络输入用于作为所述交叉网络的输入，所述第二网络输入用于作为所述深度网络的输入；可以根据所述损失更新所述第一推荐模型和所述第二特征适配网络，以得到第二推荐模型和更新后的所述第二特征适配网络。

在一种可能的实现中，所述第二特征适配网络为全连接网络、压缩奖惩网络、注意力网络、SENet或门网络。

在一种可能的实现中，所述用户属性包括如下的至少一种：性别，年龄，职业，收入，爱好，教育程度。

在一种可能的实现中，所述物品属性包括如下的至少一种：物品名称，开发者，安装包大小，品类，好评度。

参照图11，图11为本申请实施例提供的一种推荐装置1100的结构示意，所述装置1100包括：

获取模块1101，用于获取目标特征向量，所述目标特征向量为对目标用户和目标物品的属性信息或进行特征提取得到的。

关于获取模块1101的具体描述可以参照上述实施例中步骤501的描述，这里不再赘述。

数据处理模块1102，用于通过推荐模型处理所述目标特征向量，以得到推荐信息，所述推荐信息用于表示所述目标用户对所述目标物品进行选择的概率；其中，所述推荐模型包括交叉网络（cross network）、深度网络（deep network）以及目标网络，所述交叉网络包括第一交叉层（cross layer）和第二交叉层，所述深度网络包括第一深度层（deep layer）和第二深度层；所述目标网络用于对所述第一交叉层输出的第一中间输出和所述第一深度层输出的第二中间输出进行融合处理，以得到第一融合结果，所述目标网络还用于处理所述第一融合结果，以得到所述第一交叉层对应的第一权重以及所述第一深度层对应的第二权重，并将所述第一融合结果分别与所述第一权重和所述第二权重进行加权处理，以得到第一中间输入和第二中间输入；所述第一中间输入为所述第二交叉层的输入数据，所述第二中间输入为所述第二深度层的输入数据。

关于数据处理模块1102的具体描述可以参照上述实施例中步骤502的描述，这里不再赘述。

推荐模块1103，用于当所述推荐信息满足预设条件，确定向所述目标用户推荐所述目标物品。

关于推荐模块1103的具体描述可以参照上述实施例中步骤503的描述，这里不再赘述。

在一种可能的实现中，所述融合处理包括按点逐位相加（point-wise addition）、哈达玛积、拼接、基于注意力机制的池化中的一个。

一种可能的实现中，所述目标网络包括第一特征适配网络；所述第一特征适配网络为全连接网络、压缩奖惩网络（squeeze-and-excitation network）、注意力网络、SENet或门网络（gatenet）；

所述数据处理模块，具体用于：

通过所述第一特征适配网络处理所述第一融合结果。

在一种可能的实现中，所述第一中间输出包括M个第一特征向量，所述第二中间输出包括M个第二特征向量，所述第一融合结果包括M个第三特征向量，所述第一权重包括每个第一特征向量对应的第一权重值，所述第二权重包括每个第二特征向量对应的第二权重值；

所述数据处理模块，具体用于：

将所述M个第三特征向量中的每个第一特征向量与对应的第一权重值进行加权处理；

将所述M个第三特征向量中的每个第二特征向量与对应的第二权重值进行加权处理。

在一种可能的实现中，所述交叉网络还包括第三交叉层，所述深度网络还包括第三深度层；所述目标网络还用于对所述第二交叉层输出的第三中间输出和所述第二深度层输出的第四中间输出进行融合处理，以得到第二融合结果，所述目标网络还用于处理所述第二

融合结果，以得到所述第二交叉层对应的第三权重以及所述第二深度层对应的第四权重，并将所述第二融合结果分别与所述第三权重和所述第四权重进行加权处理，以得到所述第二交叉层对应的第三中间输入和所述第二深度层对应的第四中间输入；所述第三交叉层用于处理所述第三中间输入，所述第三深度层用于处理所述第四中间输入。

在一种可能的实现中，所述目标网络包括第二特征适配网络；所述第二特征适配网络为全连接网络、压缩奖惩网络（squeeze-and-excitation network）、注意力网络、SENet或门网络（gatenet）；

所述数据处理模块，具体用于：

通过所述第二特征适配网络处理所述第二融合结果。

在一种可能的实现中，所述目标特征向量包括第一网络输入和第二网络输入；

所述获取模块，具体用于：

获取初始特征向量，所述初始特征向量为对目标用户和目标物品的属性信息或进行特征提取得到的；

通过第三特征适配网络处理所述初始特征向量，以得到所述交叉网络对应的第五权重以及所述深度网络对应的第六权重，并将所述初始特征向量分别与所述第五权重和所述第六权重进行加权处理，以得到所述交叉网络对应的第一网络输入和所述深度网络对应的第二网络输入，所述第一网络输入用于作为所述交叉网络的输入，所述第二网络输入用于作为所述深度网络的输入。

在一种可能的实现中，所述第二特征适配网络为全连接网络、压缩奖惩网络、注意力网络、SENet或门网络。

在一种可能的实现中，所述用户属性包括如下的至少一种：性别，年龄，职业，收入，爱好，教育程度。

所述物品属性包括如下的至少一种：物品名称，开发者，安装包大小，品类，好评度。

参照图12，图12为本申请实施例提供的一种模型训练装置1200的结构示意，所述装置1200包括：

获取模块1201，用于获取目标特征向量，所述目标特征向量为对目标用户和目标物品的属性信息或进行特征提取得到的。

关于获取模块1201的具体描述可以参照上述实施例中步骤1001的描述，这里不再赘述。

数据处理模块1202，用于通过第一推荐模型处理所述目标特征向量，以得到推荐信息，所述推荐信息用于表示所述目标用户对所述目标物品进行选择的概率；其中，所述推荐模型包括交叉网络（cross network）、深度网络（deep network）以及目标网络，所述交叉网络包括第一交叉层（cross layer）和第二交叉层，所述深度网络包括第一深度层（deep layer）和第二深度层；所述目标网络用于对所述第一交叉层输出的第一中间输出和所述第一深度层输出的第二中间输出进行融合处理，以得到第一融合结果，所述目标网络还用于处理所述第一融合结果，以得到所述第一交叉层对应的第一权重以及所述第一深度层对应的第二权重，并将所述第一融合结果分别与所述第一权重和所述第二权重进行加权处理，以得到

第一中间输入和第二中间输入；所述第一中间输入为所述第二交叉层的输入数据，所述第二中间输入为所述第二深度层的输入数据；

关于数据处理模块1202的具体描述可以参照上述实施例中步骤1002的描述，这里不再赘述。

模型训练模块1203，用于根据所述目标用户对所述目标物品的真实选择结果、以及所述推荐信息，确定损失，并根据所述损失更新所述第一推荐模型，以得到第二推荐模型。

关于模型训练模块1203的具体描述可以参照上述实施例中步骤1003的描述，这里不再赘述。

在一种可能的实现中，所述融合处理包括按点逐位相加（point-wise addition）、哈达玛积、拼接、基于注意力机制的池化中的一个。

在一种可能的实现中，所述目标网络包括第一特征适配网络；所述第一特征适配网络为全连接网络、压缩奖惩网络（squeeze-and-excitation network）、注意力网络、SENet或门网络（gatenet）；

所述数据处理模块，具体用于：

通过所述第一特征适配网络处理所述第一融合结果。

在一种可能的实现中，所述第一中间输出包括M个第一特征向量，所述第二中间输出包括M个第二特征向量，所述第一融合结果包括M个第三特征向量，所述第一权重包括每个第一特征向量对应的第一权重值，所述第二权重包括每个第二特征向量对应的第二权重值；

所述数据处理模块，具体用于：

将所述M个第三特征向量中的每个第一特征向量与对应的第一权重值进行加权处理；

将所述M个第三特征向量中的每个第二特征向量与对应的第二权重值进行加权处理。

在一种可能的实现中，所述交叉网络还包括第三交叉层，所述深度网络还包括第三深度层；所述目标网络还用于对所述第二交叉层输出的第三中间输出和所述第二深度层输出的第四中间输出进行融合处理，以得到第二融合结果，所述目标网络还用于处理所述第二融合结果，以得到所述第二交叉层对应的第三权重以及所述第二深度层对应的第四权重，并将所述第二融合结果分别与所述第三权重和所述第四权重进行加权处理，以得到所述第二交叉层对应的第三中间输入和所述第二深度层对应的第四中间输入；所述第三交叉层用于处理所述第三中间输入，所述第三深度层用于处理所述第四中间输入。

在一种可能的实现中，所述目标网络包括第二特征适配网络；所述第二特征适配网络为全连接网络、压缩奖惩网络（squeeze-and-excitation network）、注意力网络、SENet或门网络（gatenet）；

所述数据处理模块，具体用于：

通过所述第二特征适配网络处理所述第二融合结果。

在一种可能的实现中，所述目标特征向量包括第一网络输入和第二网络输入；

所述获取模块，具体用于：

获取初始特征向量，所述初始特征向量为对目标用户和目标物品的属性信息或进行特

征提取得到的；

通过第三特征适配网络处理所述初始特征向量，以得到所述交叉网络对应的第五权重以及所述深度网络对应的第六权重，并将所述初始特征向量分别与所述第五权重和所述第六权重进行加权处理，以得到所述交叉网络对应的第一网络输入和所述深度网络对应的第二网络输入，所述第一网络输入用于作为所述交叉网络的输入，所述第二网络输入用于作为所述深度网络的输入；

所述根据所述损失更新所述第一推荐模型，以得到第二推荐模型，包括：

根据所述损失更新所述第一推荐模型和所述第二特征适配网络，以得到第二推荐模型和更新后的所述第二特征适配网络。

在一种可能的实现中，所述第二特征适配网络为全连接网络、压缩奖惩网络、注意力网络、SENet或门网络。

在一种可能的实现中，所述用户属性包括如下的至少一种：性别，年龄，职业，收入，爱好，教育程度。

在一种可能的实现中，所述物品属性包括如下的至少一种：物品名称，开发者，安装包大小，品类，好评度。

接下来介绍本申请实施例提供的一种执行设备，请参阅图 13，图 13 为本申请实施例提供的执行设备的一种结构示意图，执行设备 1300 具体可以表现为手机、平板、笔记本电脑、智能穿戴设备、服务器等，此处不做限定。其中，执行设备 1300 上可以部署有图 11 对应实施例中所描述的推荐装置，用于实现图 10 对应实施例中推荐方法的功能。具体的，执行设备 1300 包括：接收器 1301、发射器 1302、处理器 1303 和存储器 1304 (其中执行设备 1300 中的处理器 1303 的数量可以一个或多个)，其中，处理器 1303 可以包括应用处理器 13031 和通信处理器 13032。在本申请的一些实施例中，接收器 1301、发射器 1302、处理器 1303 和存储器 1304 可通过总线或其它方式连接。

存储器 1304 可以包括只读存储器和随机存取存储器，并向处理器 1303 提供指令和数据。存储器 1304 的一部分还可以包括非易失性随机存取存储器（non-volatile random access memory，NVRAM）。存储器 1304 存储有处理器和操作指令、可执行模块或者数据结构，或者它们的子集，或者它们的扩展集，其中，操作指令可包括各种操作指令，用于实现各种操作。

处理器 1303 控制执行设备的操作。具体的应用中，执行设备的各个组件通过总线系统耦合在一起，其中总线系统除包括数据总线之外，还可以包括电源总线、控制总线和状态信号总线等。但是为了清楚说明起见，在图中将各种总线都称为总线系统。

上述本申请实施例揭示的方法可以应用于处理器 1303 中，或者由处理器 1303 实现。处理器 1303 可以是一种集成电路芯片，具有信号的处理能力。在实现过程中，上述方法的各步骤可以通过处理器 1303 中的硬件的集成逻辑电路或者软件形式的指令完成。上述的处理器 1303 可以是通用处理器、数字信号处理器（digital signal processing，DSP）、微处理器或微控制器、以及视觉处理器（vision processing unit，VPU）、张量处理器（tensor processing

unit，TPU）等适用于 AI 运算的处理器，还可进一步包括专用集成电路（application specific integrated circuit，ASIC）、现场可编程门阵列（field-programmable gate array，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。该处理器 1303 可以实现或者执行本申请实施例中的公开的各方法、步骤及逻辑框图。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。结合本申请实施例所公开的方法的步骤可以直接体现为硬件译码处理器执行完成，或者用译码处理器中的硬件及软件模块组合执行完成。软件模块可以位于随机存储器，闪存、只读存储器，可编程只读存储器或者电可擦写可编程存储器、寄存器等本领域成熟的存储介质中。该存储介质位于存储器 1304，处理器 1303 读取存储器 1304 中的信息，结合其硬件完成上述实施例中步骤 501 至步骤 503 的步骤。

接收器 1301 可用于接收输入的数字或字符信息，以及产生与执行设备的相关设置以及功能控制有关的信号输入。发射器 1302 可用于通过第一接口输出数字或字符信息；发射器 1302 还可用于通过第一接口向磁盘组发送指令，以修改磁盘组中的数据；发射器 1302 还可以包括显示屏等显示设备。

本申请实施例还提供了一种训练设备，请参阅图 14，图 14 是本申请实施例提供的训练设备一种结构示意图，具体的，训练设备 1400 由一个或多个服务器实现，训练设备 1400 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中央处理器（central processing units，CPU）1414（例如，一个或一个以上处理器）和存储器 1432，一个或一个以上存储应用程序 1442 或数据 1444 的存储介质 1430（例如一个或一个以上海量存储设备）。其中，存储器 1432 和存储介质 1430 可以是短暂存储或持久存储。存储在存储介质 1430 的程序可以包括一个或一个以上模块（图示没标出），每个模块可以包括对训练设备中的一系列指令操作。更进一步地，中央处理器 1414 可以设置为与存储介质 1430 通信，在训练设备 1400 上执行存储介质 1430 中的一系列指令操作。

训练设备 1400 还可以包括一个或一个以上电源 1426，一个或一个以上有线或无线网络接口 1450，一个或一个以上输入输出接口 1458；或，一个或一个以上操作系统 1441，例如 Windows ServerTM，Mac OS XTM，UnixTM, LinuxTM，FreeBSDTM 等等。

具体的，训练设备可以进行上述实施例中步骤 1001 至步骤 1003 的步骤。

本申请实施例中还提供一种包括计算机程序产品，当其在计算机上运行时，使得计算机执行如前述执行设备所执行的步骤，或者，使得计算机执行如前述训练设备所执行的步骤。

本申请实施例中还提供一种计算机可读存储介质，该计算机可读存储介质中存储有用于进行信号处理的程序，当其在计算机上运行时，使得计算机执行如前述执行设备所执行的步骤，或者，使得计算机执行如前述训练设备所执行的步骤。

本申请实施例提供的执行设备、训练设备或终端设备具体可以为芯片，芯片包括：处理单元和通信单元，所述处理单元例如可以是处理器，所述通信单元例如可以是输入/输出接口、管脚或电路等。该处理单元可执行存储单元存储的计算机执行指令，以使执行设备

内的芯片执行上述实施例描述的数据处理方法，或者，以使训练设备内的芯片执行上述实施例描述的数据处理方法。可选地，所述存储单元为所述芯片内的存储单元，如寄存器、缓存等，所述存储单元还可以是所述无线接入设备端内的位于所述芯片外部的存储单元，如只读存储器（read-only memory，ROM)或可存储静态信息和指令的其他类型的静态存储设备，随机存取存储器（random access memory，RAM)等。

具体的，请参阅图 15，图 15 为本申请实施例提供的芯片的一种结构示意图，所述芯片可以表现为神经网络处理器 NPU1500，NPU 1500 作为协处理器挂载到主 CPU（Host CPU）上，由 Host CPU 分配任务。NPU 的核心部分为运算电路 1503，通过控制器 1504 控制运算电路 1503 提取存储器中的矩阵数据并进行乘法运算。

NPU 1500 可以通过内部的各个器件之间的相互配合，来实现图 4 所描述的实施例中提供的信息推荐方法以及图 10 所描述的实施例中提供的模型训练方法。

更具体的，在一些实现中，NPU 1500 中的运算电路 1503 内部包括多个处理单元（Process Engine, PE）。在一些实现中，运算电路 1503 是二维脉动阵列。运算电路 1503 还可以是一维脉动阵列或者能够执行例如乘法和加法这样的数学运算的其它电子线路。在一些实现中，运算电路 1503 是通用的矩阵处理器。

举例来说，假设有输入矩阵 A，权重矩阵 B，输出矩阵 C。运算电路从权重存储器 1502 中取矩阵 B 相应的数据，并缓存在运算电路中每一个 PE 上。运算电路从输入存储器 1501 中取矩阵 A 数据与矩阵 B 进行矩阵运算，得到的矩阵的部分结果或最终结果，保存在累加器（accumulator）1508 中。

统一存储器 1506 用于存放输入数据以及输出数据。权重数据直接通过存储单元访问控制器（Direct Memory Access Controller，DMAC）1505，DMAC 被搬运到权重存储器 1502 中。输入数据也通过 DMAC 被搬运到统一存储器 1506 中。

BIU 为 Bus Interface Unit 即，总线接口单元 1510，用于 AXI 总线与 DMAC 和取指存储器（Instruction Fetch Buffer，IFB）1509 的交互。

总线接口单元 1510（Bus Interface Unit，简称 BIU），用于取指存储器 1509 从外部存储器获取指令，还用于存储单元访问控制器 1505 从外部存储器获取输入矩阵 A 或者权重矩阵 B 的原数据。

DMAC 主要用于将外部存储器 DDR 中的输入数据搬运到统一存储器 1506 或将权重数据搬运到权重存储器 1502 中或将输入数据数据搬运到输入存储器 1501 中。

向量计算单元 1507 包括多个运算处理单元，在需要的情况下，对运算电路 1503 的输出做进一步处理，如向量乘，向量加，指数运算，对数运算，大小比较等等。主要用于神经网络中非卷积/全连接层网络计算，如 Batch Normalization(批归一化)，像素级求和，对特征平面进行上采样等。

在一些实现中，向量计算单元 1507 能将经处理的输出的向量存储到统一存储器 1506。例如，向量计算单元 1507 可以将线性函数；或，非线性函数应用到运算电路 1503 的输出，例如对卷积层提取的特征平面进行线性插值，再例如累加值的向量，用以生成激活值。在

一些实现中，向量计算单元 1507 生成归一化的值、像素级求和的值，或二者均有。在一些实现中，处理过的输出的向量能够用作到运算电路 1503 的激活输入，例如用于在神经网络中的后续层中的使用。

控制器 1504 连接的取指存储器（instruction fetch buffer）1509，用于存储控制器 1504 使用的指令；

统一存储器 1506，输入存储器 1501，权重存储器 1502 以及取指存储器 1509 均为 On-Chip 存储器。外部存储器私有于该 NPU 硬件架构。

其中，上述任一处提到的处理器，可以是一个通用中央处理器，微处理器，ASIC，或一个或多个用于控制上述程序执行的集成电路。

另外需说明的是，以上所描述的装置实施例仅仅是示意性的，其中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模块来实现本实施例方案的目的。另外，本申请提供的装置实施例附图中，模块之间的连接关系表示它们之间具有通信连接，具体可以实现为一条或多条通信总线或信号线。

通过以上的实施方式的描述，所属领域的技术人员可以清楚地了解到本申请可借助软件加必需的通用硬件的方式来实现，当然也可以通过专用硬件包括专用集成电路、专用 CPU、专用存储器、专用元器件等来实现。一般情况下，凡由计算机程序完成的功能都可以很容易地用相应的硬件来实现，而且，用来实现同一功能的具体硬件结构也可以是多种多样的，例如模拟电路、数字电路或专用电路等。但是，对本申请而言更多情况下软件程序实现是更佳的实施方式。基于这样的理解，本申请的技术方案本质上或者说对现有技术做出贡献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可读取的存储介质中，如计算机的软盘、U 盘、移动硬盘、ROM、RAM、磁碟或者光盘等，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，训练设备，或者网络设备等）执行本申请各个实施例所述的方法。

在上述实施例中，可以全部或部分地通过软件、硬件、固件或者其任意组合来实现。当使用软件实现时，可以全部或部分地以计算机程序产品的形式实现。

所述计算机程序产品包括一个或多个计算机指令。在计算机上加载和执行所述计算机程序指令时，全部或部分地产生按照本申请实施例所述的流程或功能。所述计算机可以是通用计算机、专用计算机、计算机网络、或者其他可编程装置。所述计算机指令可以存储在计算机可读存储介质中，或者从一个计算机可读存储介质向另一计算机可读存储介质传输，例如，所述计算机指令可以从一个网站站点、计算机、训练设备或数据中心通过有线（例如同轴电缆、光纤、数字用户线（DSL））或无线（例如红外、无线、微波等）方式向另一个网站站点、计算机、训练设备或数据中心进行传输。所述计算机可读存储介质可以是计算机能够存储的任何可用介质或者是包含一个或多个可用介质集成的训练设备、数据中心等数据存储设备。所述可用介质可以是磁性介质，（例如，软盘、硬盘、磁带）、光介质（例如，DVD）、或者半导体介质（例如固态硬盘（Solid State Disk，SSD））等。

权 利 要 求

1.一种推荐方法，其特征在于，所述方法包括：

获取目标特征向量，所述目标特征向量为对目标用户和目标物品的属性信息或进行特征提取得到的；

通过推荐模型处理所述目标特征向量，以得到推荐信息，所述推荐信息用于表示所述目标用户对所述目标物品进行选择的概率；其中，所述推荐模型包括交叉网络（cross network）、深度网络（deep network）以及目标网络，所述交叉网络包括第一交叉层（cross layer）和第二交叉层，所述深度网络包括第一深度层（deep layer）和第二深度层；所述目标网络用于对所述第一交叉层输出的第一中间输出和所述第一深度层输出的第二中间输出进行融合处理，以得到第一融合结果，所述目标网络还用于处理所述第一融合结果，以得到所述第一交叉层对应的第一权重以及所述第一深度层对应的第二权重，并将所述第一融合结果分别与所述第一权重和所述第二权重进行加权处理，以得到第一中间输入和第二中间输入；所述第一中间输入为所述第二交叉层的输入数据，所述第二中间输入为所述第二深度层的输入数据；

当所述推荐信息满足预设条件，确定向所述目标用户推荐所述目标物品。

2.根据权利要求1所述的方法，其特征在于，所述融合处理包括按点逐位相加（point-wise addition）、哈达玛积、拼接、基于注意力机制的池化中的一个。

3.根据权利要求1或2所述的方法，其特征在于，所述目标网络包括第一特征适配网络；所述第一特征适配网络为全连接网络、压缩奖惩网络（squeeze-and-excitation network）、注意力网络、SENet或门网络（gatenet）；

所述处理所述第一融合结果，包括：

通过所述第一特征适配网络处理所述第一融合结果。

4.根据权利要求1至3任一所述的方法，其特征在于，所述第一中间输出包括M个第一特征向量，所述第二中间输出包括M个第二特征向量，所述第一融合结果包括M个第三特征向量，所述第一权重包括每个第一特征向量对应的第一权重值，所述第二权重包括每个第二特征向量对应的第二权重值；

所述将所述第一融合结果分别与所述第一权重和所述第二权重进行加权处理，包括：

将所述M个第三特征向量中的每个第一特征向量与对应的第一权重值进行加权处理；

将所述M个第三特征向量中的每个第二特征向量与对应的第二权重值进行加权处理。

5. 根据权利要求1至4任一所述的方法，其特征在于，

所述交叉网络还包括第三交叉层，所述深度网络还包括第三深度层；所述目标网络还用于对所述第二交叉层输出的第三中间输出和所述第二深度层输出的第四中间输出进行融合处理，以得到第二融合结果，所述目标网络还用于处理所述第二融合结果，以得到所述

第二交叉层对应的第三权重以及所述第二深度层对应的第四权重，并将所述第二融合结果分别与所述第三权重和所述第四权重进行加权处理，以得到所述第二交叉层对应的第三中间输入和所述第二深度层对应的第四中间输入；所述第三交叉层用于处理所述第三中间输入，所述第三深度层用于处理所述第四中间输入。

6.根据权利要求5所述的方法，其特征在于，所述目标网络包括第二特征适配网络；所述第二特征适配网络为全连接网络、压缩奖惩网络（squeeze-and-excitation network）、注意力网络、SENet或门网络（gatenet）；

所述处理所述第二融合结果，包括：

通过所述第二特征适配网络处理所述第二融合结果。

7. 根据权利要求1至6任一所述的方法，其特征在于，所述目标特征向量包括第一网络输入和第二网络输入；

所述获取目标特征向量，包括：

获取初始特征向量，所述初始特征向量为对目标用户和目标物品的属性信息或进行特征提取得到的；

通过第三特征适配网络处理所述初始特征向量，以得到所述交叉网络对应的第五权重以及所述深度网络对应的第六权重，并将所述初始特征向量分别与所述第五权重和所述第六权重进行加权处理，以得到所述交叉网络对应的第一网络输入和所述深度网络对应的第二网络输入，所述第一网络输入用于作为所述交叉网络的输入，所述第二网络输入用于作为所述深度网络的输入。

8.根据权利要求7所述的方法，其特征在于，所述第三特征适配网络为全连接网络、压缩奖惩网络、注意力网络、SENet或门网络。

9.根据权利要求1至8任一所述的方法，其特征在于，所述用户属性包括如下的至少一种：性别，年龄，职业，收入，爱好，教育程度。

10. 根据权利要求1至9任一所述的方法，其特征在于，所述物品属性包括如下的至少一种：物品名称，开发者，安装包大小，品类，好评度。

11. 一种模型训练方法，其特征在于，所述方法包括：

获取目标特征向量，所述目标特征向量为对目标用户和目标物品的属性信息或进行特征提取得到的；

通过第一推荐模型处理所述目标特征向量，以得到推荐信息，所述推荐信息用于表示所述目标用户对所述目标物品进行选择的概率；其中，所述推荐模型包括交叉网络（cross network）、深度网络（deep network）以及目标网络，所述交叉网络包括第一交叉层（cross

layer）和第二交叉层，所述深度网络包括第一深度层（deep layer）和第二深度层；所述目标网络用于对所述第一交叉层输出的第一中间输出和所述第一深度层输出的第二中间输出进行融合处理，以得到第一融合结果，所述目标网络还用于处理所述第一融合结果，以得到所述第一交叉层对应的第一权重以及所述第一深度层对应的第二权重，并将所述第一融合结果分别与所述第一权重和所述第二权重进行加权处理，以得到第一中间输入和第二中间输入；所述第一中间输入为所述第二交叉层的输入数据，所述第二中间输入为所述第二深度层的输入数据；

根据所述目标用户对所述目标物品的真实选择结果、以及所述推荐信息，确定损失，并根据所述损失更新所述第一推荐模型，以得到第二推荐模型。

12.根据权利要求11所述的方法，其特征在于，所述融合处理包括按点逐位相加（point-wise addition）、哈达玛积、拼接、基于注意力机制的池化中的一个。

13.根据权利要求11或12所述的方法，其特征在于，所述目标网络包括第一特征适配网络；所述第一特征适配网络为全连接网络、压缩奖惩网络（squeeze-and-excitation network）、注意力网络、SENet或门网络（gatenet）；

所述处理所述第一融合结果，包括：

通过所述第一特征适配网络处理所述第一融合结果。

14.根据权利要求11至13任一所述的方法，其特征在于，所述第一中间输出包括M个第一特征向量，所述第二中间输出包括M个第二特征向量，所述第一融合结果包括M个第三特征向量，所述第一权重包括每个第一特征向量对应的第一权重值，所述第二权重包括每个第二特征向量对应的第二权重值；

所述将所述第一融合结果分别与所述第一权重和所述第二权重进行加权处理，包括：

将所述M个第三特征向量中的每个第一特征向量与对应的第一权重值进行加权处理；

将所述M个第三特征向量中的每个第二特征向量与对应的第二权重值进行加权处理。

15. 根据权利要求11至14任一所述的方法，其特征在于，

所述交叉网络还包括第三交叉层，所述深度网络还包括第三深度层；所述目标网络还用于对所述第二交叉层输出的第三中间输出和所述第二深度层输出的第四中间输出进行融合处理，以得到第二融合结果，所述目标网络还用于处理所述第二融合结果，以得到所述第二交叉层对应的第三权重以及所述第二深度层对应的第四权重，并将所述第二融合结果分别与所述第三权重和所述第四权重进行加权处理，以得到所述第二交叉层对应的第三中间输入和所述第二深度层对应的第四中间输入；所述第三交叉层用于处理所述第三中间输入，所述第三深度层用于处理所述第四中间输入。

16.根据权利要求15所述的方法，其特征在于，所述目标网络包括第二特征适配网络；

所述第二特征适配网络为全连接网络、压缩奖惩网络（squeeze-and-excitation network）、注意力网络、SENet或门网络（gatenet）；

所述处理所述第二融合结果，包括：

通过所述第二特征适配网络处理所述第二融合结果。

17. 根据权利要求11至16任一所述的方法，其特征在于，所述目标特征向量包括第一网络输入和第二网络输入；

所述获取目标特征向量，包括：

获取初始特征向量，所述初始特征向量为对目标用户和目标物品的属性信息或进行特征提取得到的；

通过第三特征适配网络处理所述初始特征向量，以得到所述交叉网络对应的第五权重以及所述深度网络对应的第六权重，并将所述初始特征向量分别与所述第五权重和所述第六权重进行加权处理，以得到所述交叉网络对应的第一网络输入和所述深度网络对应的第二网络输入，所述第一网络输入用于作为所述交叉网络的输入，所述第二网络输入用于作为所述深度网络的输入；

所述根据所述损失更新所述第一推荐模型，以得到第二推荐模型，包括：

根据所述损失更新所述第一推荐模型和所述第二特征适配网络，以得到第二推荐模型和更新后的所述第二特征适配网络。

18.一种推荐装置，其特征在于，所述装置包括：

获取模块，用于获取目标特征向量，所述目标特征向量为对目标用户和目标物品的属性信息或进行特征提取得到的；

数据处理模块，用于通过推荐模型处理所述目标特征向量，以得到推荐信息，所述推荐信息用于表示所述目标用户对所述目标物品进行选择的概率；其中，所述推荐模型包括交叉网络（cross network）、深度网络（deep network）以及目标网络，所述交叉网络包括第一交叉层（cross layer）和第二交叉层，所述深度网络包括第一深度层（deep layer）和第二深度层；所述目标网络用于对所述第一交叉层输出的第一中间输出和所述第一深度层输出的第二中间输出进行融合处理，以得到第一融合结果，所述目标网络还用于处理所述第一融合结果，以得到所述第一交叉层对应的第一权重以及所述第一深度层对应的第二权重，并将所述第一融合结果分别与所述第一权重和所述第二权重进行加权处理，以得到第一中间输入和第二中间输入；所述第一中间输入为所述第二交叉层的输入数据，所述第二中间输入为所述第二深度层的输入数据；

推荐模块，用于当所述推荐信息满足预设条件，确定向所述目标用户推荐所述目标物品。

19.根据权利要求18所述的装置，其特征在于，所述融合处理包括按点逐位相加（point-wise addition）、哈达玛积、拼接、基于注意力机制的池化中的一个。

20.根据权利要求18或19所述的装置，其特征在于，所述目标网络包括第一特征适配网络；所述第一特征适配网络为全连接网络、压缩奖惩网络（squeeze-and-excitation network）、注意力网络、SENet或门网络（gatenet）；

所述数据处理模块，具体用于：

通过所述第一特征适配网络处理所述第一融合结果。

21.根据权利要求18至20任一所述的装置，其特征在于，所述第一中间输出包括M个第一特征向量，所述第二中间输出包括M个第二特征向量，所述第一融合结果包括M个第三特征向量，所述第一权重包括每个第一特征向量对应的第一权重值，所述第二权重包括每个第二特征向量对应的第二权重值；

所述数据处理模块，具体用于：

将所述M个第三特征向量中的每个第一特征向量与对应的第一权重值进行加权处理；

将所述M个第三特征向量中的每个第二特征向量与对应的第二权重值进行加权处理。

22. 根据权利要求18至21任一所述的装置，其特征在于，

所述交叉网络还包括第三交叉层，所述深度网络还包括第三深度层；所述目标网络还用于对所述第二交叉层输出的第三中间输出和所述第二深度层输出的第四中间输出进行融合处理，以得到第二融合结果，所述目标网络还用于处理所述第二融合结果，以得到所述第二交叉层对应的第三权重以及所述第二深度层对应的第四权重，并将所述第二融合结果分别与所述第三权重和所述第四权重进行加权处理，以得到所述第二交叉层对应的第三中间输入和所述第二深度层对应的第四中间输入；所述第三交叉层用于处理所述第三中间输入，所述第三深度层用于处理所述第四中间输入。

23. 根据权利要求18至22任一所述的装置，其特征在于，所述目标特征向量包括第一网络输入和第二网络输入；

所述获取模块，具体用于：

获取初始特征向量，所述初始特征向量为对目标用户和目标物品的属性信息或进行特征提取得到的；

通过第三特征适配网络处理所述初始特征向量，以得到所述交叉网络对应的第五权重以及所述深度网络对应的第六权重，并将所述初始特征向量分别与所述第五权重和所述第六权重进行加权处理，以得到所述交叉网络对应的第一网络输入和所述深度网络对应的第二网络输入，所述第一网络输入用于作为所述交叉网络的输入，所述第二网络输入用于作为所述深度网络的输入。

24. 一种模型训练装置，其特征在于，所述装置包括：

获取模块，用于获取目标特征向量，所述目标特征向量为对目标用户和目标物品的属

性信息或进行特征提取得到的；

数据处理模块，用于通过第一推荐模型处理所述目标特征向量，以得到推荐信息，所述推荐信息用于表示所述目标用户对所述目标物品进行选择的概率；其中，所述推荐模型包括交叉网络（cross network）、深度网络（deep network）以及目标网络，所述交叉网络包括第一交叉层（cross layer）和第二交叉层，所述深度网络包括第一深度层（deep layer）和第二深度层；所述目标网络用于对所述第一交叉层输出的第一中间输出和所述第一深度层输出的第二中间输出进行融合处理，以得到第一融合结果，所述目标网络还用于处理所述第一融合结果，以得到所述第一交叉层对应的第一权重以及所述第一深度层对应的第二权重，并将所述第一融合结果分别与所述第一权重和所述第二权重进行加权处理，以得到第一中间输入和第二中间输入；所述第一中间输入为所述第二交叉层的输入数据，所述第二中间输入为所述第二深度层的输入数据；

模型训练模块，用于根据所述目标用户对所述目标物品的真实选择结果、以及所述推荐信息，确定损失，并根据所述损失更新所述第一推荐模型，以得到第二推荐模型。

25.根据权利要求24所述的装置，其特征在于，所述融合处理包括按点逐位相加（point-wise addition）、哈达玛积、拼接、基于注意力机制的池化中的一个。

26.根据权利要求24或25所述的装置，其特征在于，所述目标网络包括第一特征适配网络；所述第一特征适配网络为全连接网络、压缩奖惩网络（squeeze-and-excitation network）、注意力网络、SENet或门网络（gatenet）；

所述数据处理模块，具体用于：

通过所述第一特征适配网络处理所述第一融合结果。

27.根据权利要求24至26任一所述的装置，其特征在于，所述第一中间输出包括M个第一特征向量，所述第二中间输出包括M个第二特征向量，所述第一融合结果包括M个第三特征向量，所述第一权重包括每个第一特征向量对应的第一权重值，所述第二权重包括每个第二特征向量对应的第二权重值；

所述数据处理模块，具体用于：

将所述M个第三特征向量中的每个第一特征向量与对应的第一权重值进行加权处理；

将所述M个第三特征向量中的每个第二特征向量与对应的第二权重值进行加权处理。

28. 根据权利要求24至27任一所述的装置，其特征在于，所述目标特征向量包括第一网络输入和第二网络输入；

所述获取模块，具体用于：

获取初始特征向量，所述初始特征向量为对目标用户和目标物品的属性信息或进行特征提取得到的；

通过第三特征适配网络处理所述初始特征向量，以得到所述交叉网络对应的第五权重

以及所述深度网络对应的第六权重，并将所述初始特征向量分别与所述第五权重和所述第六权重进行加权处理，以得到所述交叉网络对应的第一网络输入和所述深度网络对应的第二网络输入，所述第一网络输入用于作为所述交叉网络的输入，所述第二网络输入用于作为所述深度网络的输入；

所述模型训练模块，具体用于：

根据所述损失更新所述第一推荐模型和所述第二特征适配网络，以得到第二推荐模型和更新后的所述第二特征适配网络。

29.一种计算设备，其特征在于，所述计算设备包括存储器和处理器；所述存储器存储有代码，所述处理器被配置为获取所述代码，并执行如权利要求1至17任一所述的方法。

30.一种计算机存储介质，其特征在于，所述计算机存储介质存储有一个或多个指令，所述指令在由一个或多个计算机执行时使得所述一个或多个计算机实施权利要求1至17任一所述的方法。

31.一种计算机程序产品，包括代码，其特征在于，在所述代码被执行时用于实现如权利要求 1 至 17 任一所述的方法。