



(12) 发明专利

(10) 授权公告号 CN 113626719 B

(45) 授权公告日 2022. 02. 08

(21) 申请号 202111184748.8

G06N 3/04 (2006.01)

(22) 申请日 2021.10.12

G06N 3/08 (2006.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 113626719 A

(56) 对比文件

CN 112232546 A, 2021.01.15

CN 108763362 A, 2018.11.06

US 2020090250 A1, 2020.03.19

US 2020233873 A1, 2020.07.23

CN 112699305 A, 2021.04.23

CN 112699305 A, 2021.04.23

(43) 申请公布日 2021.11.09

(73) 专利权人 腾讯科技(深圳)有限公司

地址 518057 广东省深圳市南山区高新区

科技中一路腾讯大厦35层

审查员 罗晓彤

(72) 发明人 马骊 赵忠 梁瀚明 赵光耀

傅妍玫 户维波 何新昇 吴铭津

(74) 专利代理机构 北京派特恩知识产权代理有

限公司 11270

代理人 刘星雨 张颖玲

(51) Int. Cl.

G06F 16/9535 (2019.01)

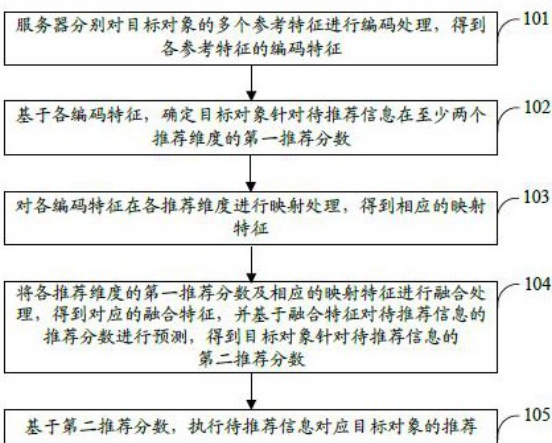
权利要求书3页 说明书21页 附图9页

(54) 发明名称

信息推荐方法、装置、设备、存储介质及计算机程序产品

(57) 摘要

本申请提供了一种信息推荐方法、装置、设备、存储介质及程序产品,应用于车联网领域以及人工智能技术领域;方法包括:分别对目标对象的多个参考特征进行编码处理,得到各参考特征的编码特征;基于各编码特征,确定目标对象针对待推荐信息在至少两个推荐维度的第一推荐分数;对各编码特征在各推荐维度进行映射处理,得到相应的映射特征;将各推荐维度的第一推荐分数及相应的映射特征进行融合处理,得到对应的融合特征,并基于融合特征对待推荐信息的推荐分数进行预测,得到目标对象针对待推荐信息的第二推荐分数;基于第二推荐分数,执行待推荐信息对应目标对象的推荐。



1. 一种信息推荐方法,其特征在于,所述方法包括:

分别对目标对象的多个参考特征进行编码处理,得到各所述参考特征的编码特征;

分别将所述编码特征的交叉特征、及待推荐信息在至少两个对应推荐维度的拟合特征进行拼接处理,得到对应各所述推荐维度的拼接特征,其中,拟合特征为用于表征所有推荐维度中任意两个推荐维度之间的拟合自由度的特征;

对各所述推荐维度的拼接特征进行映射处理,得到各所述拼接特征的映射特征,通过激活函数对各所述拼接特征的映射特征进行偏置处理,得到对应各所述推荐维度的第一推荐分数;

对各所述编码特征在各所述推荐维度进行映射处理,得到各所述编码特征相应的映射特征,各所述编码特征的映射特征用于表征在相应的所述推荐维度所述第一推荐分数的融合权重;

将各所述推荐维度的第一推荐分数及相应的各所述编码特征的映射特征进行融合处理,得到对应的融合特征,并基于所述融合特征对所述待推荐信息的推荐分数进行预测,得到所述目标对象针对所述待推荐信息的第二推荐分数;

基于所述第二推荐分数,执行所述待推荐信息对应所述目标对象的推荐。

2. 如权利要求1所述的方法,其特征在于,所述参考特征包括连续型特征及离散型特征,所述分别对目标对象的多个参考特征进行编码处理,得到各所述参考特征的编码特征,包括:

对所述连续型特征进行离散化处理,得到所述连续型特征的离散特征;

对所述连续型特征的离散特征进行编码处理,得到所述连续型特征的编码特征;

对所述离散型特征进行编码处理,得到所述离散型特征的编码特征;

将所述连续型特征的编码特征及所述离散型特征的编码特征确定为各所述参考特征的编码特征。

3. 如权利要求1所述的方法,其特征在于,在进行所述拼接处理之前,所述方法还包括:对各所述编码特征进行特征交叉,得到对应的交叉特征;

其中,所述对各所述编码特征进行特征交叉,得到对应的交叉特征,包括:

对所述多个参考特征进行一阶特征处理,得到对应的一阶特征;

对多个所述编码特征中任意两个编码特征进行二阶交叉处理,得到所述任意两个编码特征的二阶交叉特征;

将得到的所述一阶特征及所述二阶交叉特征进行融合处理,得到对应的交叉特征。

4. 如权利要求1所述的方法,其特征在于,在进行所述拼接处理之前,所述方法还包括:基于各所述编码特征,对所述待推荐信息在至少两个推荐维度的拟合性进行预测,得到对应各推荐维度的拟合特征;

其中,所述基于各所述编码特征,对所述待推荐信息在至少两个推荐维度的拟合性进行预测,得到对应各推荐维度的拟合特征,包括:通过多门控混合专家网络中对应各所述推荐维度的专家网络,对各所述编码特征进行全连接处理,得到对应的第一隐层特征,并对所述第一隐层特征进行映射处理,得到对应各所述专家网络的映射特征;

通过所述多门控混合专家网络中对应各所述推荐维度的门控网络,对各所述编码特征进行全连接处理,得到对应的第二隐层特征,并对所述第二隐层特征进行映射处理,得到对

应各所述专家网络的权重特征；

基于对应各所述专家网络的权重特征，对对应各所述专家网络的映射特征进行加权求和处理，得到对应各所述推荐维度的拟合特征。

5. 如权利要求1所述的方法，其特征在于，所述对各所述编码特征在各所述推荐维度进行映射处理，得到各所述编码特征相应的映射特征，包括：

对各所述推荐维度的第一推荐分数进行横向拼接处理，得到对应的平铺向量；

对各所述编码特征进行全连接处理，得到对应的隐层特征；

对所述隐层特征进行映射处理，得到与所述平铺向量维度相同的各所述编码特征的映射特征。

6. 如权利要求1所述的方法，其特征在于，所述将各所述推荐维度的第一推荐分数及相应的各所述编码特征的映射特征进行融合处理，得到对应的融合特征，包括：

获取由各所述推荐维度的第一推荐分数构成的分数矩阵，以及各所述推荐维度对应的各所述编码特征的映射特征构成的映射矩阵；

将所述分数矩阵与所述映射矩阵进行元素积计算，得到对应的融合特征。

7. 如权利要求1所述的方法，其特征在于，所述基于所述融合特征对所述待推荐信息的推荐分数进行预测，得到所述目标对象针对所述待推荐信息的第二推荐分数，包括：

对所述融合特征进行映射处理，得到所述融合特征对应的映射特征；

基于所述融合特征对应的映射特征对所述待推荐信息的推荐分数进行预测，得到所述目标对象针对所述待推荐信息的第二推荐分数。

8. 如权利要求1至7任一项所述的方法，其特征在于，所述信息推荐方法是通过调用分数预测模型实现的，所述分数预测模型包括：特征编码层、第一推荐分数预测层、特征映射层和第二推荐分数预测层；所述方法还包括：

通过所述特征编码层，分别对训练样本的多个参考特征进行编码处理，得到对应各所述参考特征的编码特征，所述训练样本携带对象样本针对信息样本在至少两个推荐维度的第一标签，以及所述对象样本针对所述信息样本的第二标签；

通过所述第一推荐分数预测层，基于各所述编码特征，确定所述对象样本针对所述信息样本在至少两个推荐维度的第一预测结果；

通过所述特征映射层，对各所述编码特征在各所述推荐维度进行映射处理，得到各所述编码特征的映射特征，各所述编码特征的映射特征用于表征在相应的所述推荐维度所述第一推荐分数的融合权重；

通过所述第二推荐分数预测层，将各所述推荐维度的第一推荐分数及相应的各所述编码特征的映射特征进行融合处理，得到对应的融合特征，并基于所述融合特征，对所述信息样本的推荐分数进行预测，得到所述对象样本针对所述信息样本的第二预测结果；

基于各所述推荐维度的第一预测结果与相应的第一标签，及所述第二预测结果与所述第二标签，更新所述分数预测模型的模型参数。

9. 如权利要求8所述的方法，其特征在于，所述基于各所述推荐维度的第一预测结果与相应的第一标签，及所述第二预测结果与所述第二标签，更新所述分数预测模型的模型参数，包括：

针对各所述推荐维度，基于所述第一预测结果及相应的第一标签，构造对应所述第一

推荐分数预测层的第一损失函数；

基于所述第二预测结果及所述第二标签，构造对应所述第二推荐分数预测层的第二损失函数；

将所述第二损失函数及所述第一损失函数进行加权求和，得到所述分数预测模型的第三损失函数；

基于所述第三损失函数更新所述分数预测模型的模型参数。

10. 如权利要求9所述的方法，其特征在于，所述针对各所述推荐维度，基于所述第一预测结果及相应的第一标签，构造对应所述第一推荐分数预测层的第一损失函数，包括：

基于各所述推荐维度对应的第一预测结果及相应的第一标签，构造对应各所述推荐维度的子损失函数；

确定各所述推荐维度对应的推荐权重，基于各所述推荐权重对对应各所述推荐维度的子损失函数进行加权求和，得到对应所述第一推荐分数预测层的第一损失函数。

11. 一种信息推荐装置，其特征在于，所述装置包括：

特征编码模块，用于分别对目标对象的多个参考特征进行编码处理，得到对应各所述参考特征的编码特征；

第一预测模块，用于对各所述编码特征进行特征交叉，得到对应的交叉特征；基于各所述编码特征，对待推荐信息在至少两个推荐维度的拟合性进行预测，得到对应各推荐维度的拟合特征，其中，拟合特征为用于表征所有推荐维度中任意两个推荐维度之间的拟合自由度的特征；分别将所述交叉特征及对应各所述推荐维度的拟合特征进行拼接处理，得到对应各所述推荐维度的拼接特征；对各所述推荐维度的拼接特征进行映射处理，得到各所述拼接特征的映射特征，通过激活函数对各所述拼接特征的映射特征进行偏置处理，得到对应各所述推荐维度的第一推荐分数；

特征映射模块，用于对各所述编码特征在各所述推荐维度进行映射处理，得到各所述编码特征相应的映射特征，各所述编码特征的映射特征用于表征在相应的所述推荐维度所述第一推荐分数的融合权重；

第二预测模块，用于将各所述推荐维度的第一推荐分数及相应的各所述编码特征的映射特征进行融合处理，得到对应的融合特征，并基于所述融合特征对所述待推荐信息的推荐分数进行预测，得到所述目标对象针对所述待推荐信息的第二推荐分数；

信息推荐模块，用于基于所述第二推荐分数，执行所述待推荐信息对应所述目标对象的推荐。

12. 一种电子设备，其特征在于，所述电子设备包括：

存储器，用于存储可执行指令；

处理器，用于执行所述存储器中存储的可执行指令时，实现权利要求1至10任一项所述的信息推荐方法。

13. 一种计算机可读存储介质，其特征在于，存储有可执行指令，用于被处理器执行时，实现权利要求1至10任一项所述的信息推荐方法。

14. 一种计算机程序产品，包括计算机程序或指令，其特征在于，所述计算机程序或指令被处理器执行时实现权利要求1至10任一项所述的信息推荐方法。

信息推荐方法、装置、设备、存储介质及计算机程序产品

技术领域

[0001] 本申请涉及车联网领域以及人工智能技术,尤其涉及一种信息推荐方法、装置、设备、计算机可读存储介质及计算机程序产品。

背景技术

[0002] 人工智能(AI,Artificial Intelligence)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、方法和技术及应用系统。

[0003] 人工智能技术广泛应用于推荐系统中,例如,通过推荐系统的多目标排序模型将用户感兴趣的信息推荐给合适的用户,多目标排序模型从用户对信息的点击、消费时长及互动行为等推荐维度(又称目标)对信息进行预估打分,在得到针对每个目标的得分后,将多个得分融合的方式影响着推荐系统的精度和用户体验。

[0004] 相关技术缺乏有效的融合方案无法准确预测不同用户针对信息的得分,使得无法适用于个性化推荐以提升推荐系统的推荐精度和用户体验。

发明内容

[0005] 本申请实施例提供一种信息推荐方法、装置、设备、计算机可读存储介质及计算机程序产品,能够准确预测用户针对待推荐信息的推荐分数,以提升推荐系统的推荐精度和用户体验。

[0006] 本申请实施例的技术方案是这样实现的:

[0007] 本申请实施例提供一种信息推荐方法,包括:

[0008] 分别对目标对象的多个参考特征进行编码处理,得到各所述参考特征的编码特征;

[0009] 基于各所述编码特征,确定所述目标对象针对待推荐信息在至少两个推荐维度的第一推荐分数;

[0010] 对各所述编码特征在各所述推荐维度进行映射处理,得到相应的映射特征,所述映射特征用于表征在相应的所述推荐维度所述第一推荐分数的融合权重;

[0011] 将各所述推荐维度的第一推荐分数及相应的所述映射特征进行融合处理,得到对应的融合特征,并基于所述融合特征对所述待推荐信息的推荐分数进行预测,得到所述目标对象针对所述待推荐信息的第二推荐分数;

[0012] 基于所述第二推荐分数,执行所述待推荐信息对应所述目标对象的推荐。

[0013] 本申请实施例提供一种信息推荐装置,包括:

[0014] 特征编码模块,用于分别对目标对象的多个参考特征进行编码处理,得到对应各所述参考特征的编码特征;

[0015] 第一预测模块,用于基于各所述编码特征,确定所述目标对象针对待推荐信息在至少两个推荐维度的第一推荐分数;

[0016] 特征映射模块,用于对各所述编码特征在各所述推荐维度进行映射处理,得到相应的映射特征,所述映射特征用于表征在相应的所述推荐维度所述第一推荐分数的融合权重;

[0017] 第二预测模块,用于将各所述推荐维度的第一推荐分数及相应的所述映射特征进行融合处理,得到对应的融合特征,并基于所述融合特征对所述待推荐信息的推荐分数进行预测,得到所述目标对象针对所述待推荐信息的第二推荐分数;

[0018] 信息推荐模块,用于基于所述第二推荐分数,执行所述待推荐信息对应所述目标对象的推荐。

[0019] 上述方案中,所述参考特征包括连续型特征及离散型特征,所述特征编码模块,还用于对所述连续型特征进行离散化处理,得到所述连续型特征的离散特征;

[0020] 对所述连续型特征的离散特征进行编码处理,得到所述连续型特征的编码特征;

[0021] 对所述离散型特征进行编码处理,得到所述离散型特征的编码特征;

[0022] 将所述连续型特征的编码特征及所述离散型特征的编码特征确定为各所述参考特征的编码特征。

[0023] 上述方案中,所述第一预测模块,还用于对各所述编码特征进行特征交叉,得到对应的交叉特征;

[0024] 基于各所述编码特征,对所述待推荐信息在至少两个推荐维度的拟合性进行预测,得到对应各推荐维度的拟合特征;

[0025] 分别将所述交叉特征及对应各所述推荐维度的拟合特征进行拼接处理,得到对应各所述推荐维度的拼接特征;

[0026] 基于所述拼接特征,针对所述待推荐信息在至少两个推荐维度的推荐分数进行预测,得到对应各所述推荐维度的第一推荐分数。

[0027] 上述方案中,所述第一预测模块,还用于对所述多个参考特征进行一阶特征处理,得到对应的一阶特征;

[0028] 对多个所述编码特征中任意两个编码特征进行二阶交叉处理,得到所述任意两个编码特征的二阶交叉特征;

[0029] 将得到的所述一阶特征及所述二阶交叉特征进行融合处理,得到对应的交叉特征。

[0030] 上述方案中,所述第一预测模块,还用于通过多门控混合专家网络中对应各所述推荐维度的专家网络,对各所述编码特征进行全连接处理,得到对应的第一隐层特征,并对所述第一隐层特征进行映射处理,得到对应各所述专家网络的映射特征;

[0031] 通过所述多门控混合专家网络中对应各所述推荐维度的门控网络,对各所述编码特征进行全连接处理,得到对应的第二隐层特征,并对所述第二隐层特征进行映射处理,得到对应各所述专家网络的权重特征;

[0032] 基于对应各所述专家网络的权重特征,对对应各所述专家网络的映射特征进行加权求和处理,得到对应各所述推荐维度的拟合特征。

[0033] 上述方案中,所述特征映射模块,还用于对各所述推荐维度的第一推荐分数进行横向拼接处理,得到对应的平铺向量;

[0034] 对各所述编码特征进行全连接处理,得到对应的隐层特征;

[0035] 对所述隐层特征进行映射处理,得到与所述平铺向量维度相同的映射特征。

[0036] 上述方案中,所述第二预测模块,还用于获取由各所述推荐维度的第一推荐分数构成的分数矩阵,以及各所述推荐维度对应的映射特征构成的映射矩阵;

[0037] 将所述分数矩阵与所述映射矩阵进行元素积计算,得到对应的融合特征。

[0038] 上述方案中,所述第二预测模块,还用于对所述融合特征进行映射处理,得到对应的映射特征;

[0039] 基于所述映射特征对所述待推荐信息的推荐分数进行预测,得到所述目标对象针对所述待推荐信息的第二推荐分数。

[0040] 上述方案中,所述信息推荐方法是通过调用分数预测模型实现的,所述分数预测模型包括:特征编码层、第一推荐分数预测层、特征映射层和第二推荐分数预测层;所述装置还包括:

[0041] 模型训练模块,用于通过所述特征编码层,分别对训练样本的多个参考特征进行编码处理,得到对应各所述参考特征的编码特征,所述训练样本携带对象样本针对信息样本在至少两个推荐维度的第一标签,以及所述对象样本针对所述信息样本的第二标签;

[0042] 通过所述第一推荐分数预测层,基于各所述编码特征,确定所述对象样本针对所述信息样本在至少两个推荐维度的第一预测结果;

[0043] 通过所述特征映射层,对各所述编码特征在各所述推荐维度进行映射处理,得到相应的映射特征,所述映射特征用于表征在相应的所述推荐维度所述第一推荐分数的融合权重;

[0044] 通过所述第二推荐分数预测层,将各所述推荐维度的第一推荐分数及相应的所述映射特征进行融合处理,得到对应的融合特征,并基于所述融合特征,对所述信息样本的推荐分数进行预测,得到所述对象样本针对所述信息样本的第二预测结果;

[0045] 基于各所述推荐维度的第一预测结果与相应的第一标签,及所述第二预测结果与所述第二标签,更新所述分数预测模型的模型参数。

[0046] 上述方案中,所述模型训练模块,还用于针对各所述推荐维度,基于所述第一预测结果及相应的第一标签,构造对应所述第一推荐分数预测层的第一损失函数;

[0047] 基于所述第二预测结果及所述第二标签,构造对应所述第二推荐分数预测层的第二损失函数;

[0048] 将所述第二损失函数及所述第一损失函数进行加权求和,得到所述分数预测模型的第三损失函数;

[0049] 基于所述第三损失函数更新所述分数预测模型的模型参数。

[0050] 上述方案中,所述模型训练模块,还用于基于各所述推荐维度对应的第一预测结果及相应的第一标签,构造对应各所述推荐维度的子损失函数;

[0051] 确定各所述推荐维度对应的推荐权重,基于各所述推荐权重对对应各所述推荐维度的子损失函数进行加权求和,得到对应所述第一推荐分数预测层的第一损失函数。

[0052] 本申请实施例提供一种电子设备,包括:

[0053] 存储器,用于存储可执行指令;

[0054] 处理器,用于执行所述存储器中存储的可执行指令时,实现本申请实施例提供的信息推荐方法。

[0055] 本申请实施例提供一种计算机可读存储介质,存储有可执行指令,用于引起处理器执行时,实现本申请实施例提供的信息推荐方法。

[0056] 本申请实施例提供一种计算机程序产品,包括计算机程序或指令,所述计算机程序或指令被处理器执行时实现本申请实施例提供的信息推荐方法。

[0057] 本申请实施例具有以下有益效果:

[0058] 先通过对目标对象的多个参考特征的编码特征,预测得到目标对象针对待推荐信息在多个推荐维度(如点击、时长、互动等维度)的第一推荐分数,再通过特征映射的方式将编码特征映射至各推荐维度的第一推荐分数,得到表征在相应的推荐维度第一推荐分数的融合权重,将各推荐维度的第一推荐分数及相应的映射特征进行融合处理,基于融合结果预测目标对象针对待推荐信息最终的第二推荐分数;如此,可根据不同目标对象的参考特征,采用适合相应目标对象的融合方式进行各个推荐维度的分数的融合,实现了自动根据目标对象在不同推荐维度上的倾向性得到精确的最终推荐分数的目的,能够提高最终推荐分数的预测准确性,为推荐系统提供准确的推荐参考数据,进而提升推荐精度和用户体验。

附图说明

- [0059] 图1A是本申请实施例提供的信息推荐示意图一;
- [0060] 图1B是本申请实施例提供的信息推荐示意图二;
- [0061] 图1C是本申请实施例提供的信息推荐示意图三;
- [0062] 图2是本申请实施例提供的信息推荐系统10的架构示意图;
- [0063] 图3是本申请实施例提供的用于信息推荐的电子设备500的结构示意图;
- [0064] 图4A是本申请实施例提供的信息推荐方法的流程示意图;
- [0065] 图4B为本申请实施例提供的第一推荐分数的确定示意图;
- [0066] 图4C为本申请实施例提供的拟合特征的确定示意图;
- [0067] 图5为本申请实施例提供的信息推荐示意图四;
- [0068] 图6A为本申请实施例提供的模型训练方法的流程示意图;
- [0069] 图6B为本申请实施例提供的模型参数更新方法的流程示意图;
- [0070] 图7为本申请实施例提供的信息推荐示意图五;
- [0071] 图8A为本申请实施例提供的信息推荐效果示意图一;
- [0072] 图8B为本申请实施例提供的信息推荐效果示意图二;
- [0073] 图8C为本申请实施例提供的信息推荐效果示意图三。

具体实施方式

[0074] 为了使本申请的目的、技术方案和优点更加清楚,下面将结合附图对本申请作进一步地详细描述,所描述的实施例不应视为对本申请的限制,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它实施例,都属于本申请保护的范围。

[0075] 在以下的描述中,涉及到“一些实施例”,其描述了所有可能实施例的子集,但是可以理解,“一些实施例”可以是所有可能实施例的相同子集或不同子集,并且可以在不冲突的情况下相互结合。

[0076] 在以下的描述中,所涉及的术语“第一\第二\第三”仅仅是区别类似的对象,不代

表针对对象的特定排序,可以理解地,“第一\第二\第三”在允许的情况下可以互换特定的顺序或先后次序,以使这里描述的本申请实施例能够以除了在这里图示或描述的以外的顺序实施。

[0077] 除非另有定义,本文所使用的所有的技术和科学术语与属于本申请的技术领域的技术人员通常理解的含义相同。本文中所使用的术语只是为了描述本申请实施例的目的,不是旨在限制本申请。

[0078] 对本申请实施例进行进一步详细说明之前,对本申请实施例中涉及的名词和术语进行说明,本申请实施例中涉及的名词和术语适用于如下的解释。

[0079] 1) 点击率(CTR, Click Through Rate):是指网页面上某一信息被点击的次数与被显示次数之比。

[0080] 2) 时长(Duration):是指用户对信息的消费时长,如用户阅读信息的时长。

[0081] 3) 互动:包括但不限于用户对信息的点赞、分享、收藏、转发、关注等操作。

[0082] 4) 排序:对召回的候选信息进行打分,按照分数从召回的候选信息中选取top N个信息作为推荐结果。

[0083] 5) 嵌入式表示(Embedding):是将源数据映射到另外一个空间,在深度学习中,是一种用低维稠密向量来表示数据的方法。

[0084] 6) 多层感知器(MLP, Multiple Layer Perceptron):通常也称为深度神经网络(DNN, Deep Neural Networks),即多层全连接神经网络,通过激活函数使得矩阵特征变换加入非线性性。

[0085] 7) 多门控混合专家网络(MMoE, Multi-gate Mixture-of-Experts):用于多目标学习的常用网络结构,由多个专家网络和多个门控网络构成,其中,专家网络多为DNN网络结构,专家网络用于提取不同的特征,门控网络用于分配每个专家网络的权重。

[0086] 8) 个性化特征:根据用户之前的点击、互动历史等行为数据以及相似用户的历史行为数据预测用户的需求和偏好,进而给用户推荐可能喜欢的物品,在以上推荐过程中使用到的用户或物品的统计数据或衍生的数据。

[0087] 9) 自适应:根据处理数据的数据特征自动调整处理方法和参数权重,是一种数学模型不断逼近目标的过程。

[0088] 10) 多目标融合:多个目标经过训练有多个预估打分,各预估打分根据目标重要性、业务指标需求等策略相加或相乘,或其他融合方法进行融合排序。

[0089] 在信息推荐过程中的多目标排序模型一般用于预估多个目标(即推荐维度)的得分(即推荐分数),如何把多个得分融合成单一合理的排序分,并且在业务上实现最佳的效果是比较困难的问题。业界的一般做法有如下几种:

[0090] 1、第一类:公式融合方法

[0091] 通过对各个目标的预测模型进行单独训练,然后将预测出的不同目标的得分通过公式进行融合,或者在多目标网络基础上得到多个得分,然后通过手动设计相加、相乘或更复杂的公式进行融合,如加入指数因子带入非线性。其中,最直观的方法是线性融合,线性融合的参数满足约束 $\{\theta_1, \theta_2, \dots, \theta_k | \sum \theta_i = 1\}$,融合公式为 $f(P_{\text{点击率}}, P_{\text{时长}}, P_{\text{互动}}) = \sum (\theta_1 P_{\text{点击率}} + \theta_2 P_{\text{时长}} + \theta_3 P_{\text{互动}})$,

若相乘并加入指数因子公式可以为 $f(P_{\text{点击率}}, P_{\text{时长}}, P_{\text{互动}}) = P_{\text{点击率}}^{\alpha} \cdot P_{\text{时长}}^{\beta} \cdot P_{\text{互动}}^{\gamma}$,其中, α 、 β

和 γ 为相应的指数因子。

[0092] 为了找到相对较好的参数,还需要离线搜索不同的参数组,常用的方法有网格搜索(grid-search)或者是启发式方法(如遗传算法、粒子群算法等),先为 $\theta_1, \theta_2, \dots, \theta_k$ 各自设置搜索范围,比如从0.1到2.0,以步长0.1来构造候选参数,得到所有的候选参数组合根据离线数据的效果择优;或者是设计效果评价指标,把参数直接放到在线业务中验证效果,对比多组参数的效果后不断择优。

[0093] 2、第二类:模型融合方法

[0094] 一种是基于树模型的融合方法,如基于梯度提升树(GBDT,Gradient Boosting Decision Tree)模型的融合方法,输入多个目标的得分和一些对象特征,目标是加权的组合标签,等价于将叶子结点转换为打分规则。还有一种是加入深度网络参数自动学习打分融合权重的模型融合方法,例如,参见图1A,图1A为本申请实施例提供的信息推荐示意图,在神经网络最后一层增加一层权重生成网络,将多目标问题转化成单目标问题,如在得到每个目标(如美食、酒店、旅游)的得分后,每个目标的得分乘以相应的权重 w_i 后求和得到一个最终得分,作为排序的依据。

[0095] 参见图1B,图1B为本申请实施例提供的信息推荐示意图,在推荐过程中采用线性加权计算双塔内积的方式直接得到最终得分,物品(如为视频)塔直接把各种个性化预估值拼在一起,形成一个24维向量;用户塔的顶层向量通过网络学习,产出一个24维向量;最后,对两个塔产出的向量做内积,损失函数采用加权方式进行构造。

[0096] 参见图1C,图1C为本申请实施例提供的信息推荐示意图,在信息推荐过程中,采用端到端的网络结构进行融合,将对象特征、信息特征和目标预测的概率进行横向拼接,对于网格生成(如Pointwise)形式,将对象标识、行为序列等作为原始参考特征,同时,融入目标预测概率特征,使用精排模型来学习最终的组合得分;采用特征对(如Pairwise)形式,在一次用户请求返回的多个视频之间,对每种目标都如下操作:先通过该目标的正样本和负样本构造偏序对,再使用神经网络学习偏序对的得分,对得分做sigmoid变换,最后通过交叉熵损失得到损失函数。

[0097] 上述方式至少存在以下缺点:对于第一类公式融合方法,如果是多个目标的预测模型单独训练,此种方式成本较高,多个预测模型无法共享参数促进共同训练加速一些特征的学习,而且在线服务负载压力较大,加载预测模型个数较多,计算量相对较大,资源消耗多,稳定性较差;如果增加的新目标数据比较稀疏,难以进行有效的模型训练和迭代。可见,不论是多个预测模型单独训练还是多目标网络共同训练,这种方法过于依赖人工规则,存在离线和在线的数据分布差异,离线搜参验证效果依赖在线数据的收集和效果指标的指定,多个目标的重要性难以量化;调参需要遍历很多组参数组合,耗时耗力,难以适应业务数据的实时变化,成本高,缺少个性和场景化;当目标不断增多,公式排序能力受限,无法寻到最优参数组合,可能导致业务指标变差。故上述两类方法都无法适用于全部用户,没有考虑用户的个性化水平,由于每个用户对不同目标的倾向性不同,限制模型在全部用户上实现最优效果。

[0098] 对于第二类模型融合方法,该方法的缺点是树模型表达能力有限,与排序模型分离,无法实现在线实时调整,另外,多目标转化成单目标无法考虑不同目标的学习程度,不

同目标之间相互影响可能导致某些目标训练不完全,而受到其他目标的主导,融合的网络设计也过于简单,意义不够直观,限制了树模型的表达能力,且如果树模型和精排模型独立,不是端到端的模型,在线预测就需要分两个模型,较为复杂。

[0099] 为此,本申请实施例提供一种信息推荐方法、装置、设备、计算机可读存储介质及计算机程序产品,能够准确预测用户针对信息的推荐分数,以提升推荐系统的推荐精度和用户体验。

[0100] 本申请实施例提供的信息推荐方法可以由各种电子设备实施,例如,可以由终端单独实施,也可以由服务器单独实施,也可以由终端和服务器协同实施。例如终端独自执行下文所述的信息推荐方法,或者,终端向服务器发送推荐请求,服务器根据接收的推荐请求执行信息推荐方法。

[0101] 本申请实施例提供的用于信息推荐的电子设备可以是各种类型的终端设备或服务器,其中,服务器可以是独立的物理服务器,也可以是多个物理服务器构成的服务器集群或者分布式系统,还可以是提供云计算服务的云服务器;终端可以是智能手机、平板电脑、笔记本电脑、台式计算机、智能音箱、智能手表、车载终端等,但并不局限于此。终端以及服务器可以通过有线或无线通信方式进行直接或间接地连接,本申请在此不做限制。

[0102] 以服务器为例,例如可以是部署在云端的服务器集群,向用户开放人工智能云服务(AI as a Service,AIaaS),AIaaS平台会把几类常见的AI服务进行拆分,并在云端提供独立或者打包的服务,这种服务模式类似于一个AI主题商城,所有的用户都可以通过应用程序编程接口的方式来接入使用AIaaS平台提供的一种或者多种人工智能服务。

[0103] 例如,其中的一种人工智能云服务可以为信息推荐服务,即云端的服务器封装有本申请实施例提供的信息推荐程序。用户通过终端(运行有客户端,例如即时通信客户端、直播客户端、短视频客户端、社交客户端等)调用云服务中的信息推荐服务,以使部署在云端的服务器调用封装的信息推荐程序,确定目标对象针对待推荐信息的推荐分数,并基于推荐分数执行待推荐信息对应目标对象的推荐。

[0104] 在一些实施例中,以服务器单独实施本申请实施例提供的信息推荐方法为例进行说明。服务器分别对目标对象的多个参考特征进行编码处理,得到各参考特征的编码特征;基于各编码特征,确定目标对象针对待推荐信息在至少两个推荐维度的第一推荐分数;对各编码特征在各推荐维度进行映射处理,得到相应的映射特征,映射特征用于表征在相应的推荐维度所述第一推荐分数的融合权重;将各推荐维度的第一推荐分数及相应的映射特征进行融合处理,得到对应的融合特征,并基于融合特征对待推荐信息的推荐分数进行预测,得到目标对象针对待推荐信息的第二推荐分数;基于第二推荐分数,执行待推荐信息对应目标对象的推荐。

[0105] 在一些实施例中,以服务器和终端协同实施本申请实施例提供的信息推荐方法为例进行说明。参见图2,图2是本申请实施例提供的信息推荐系统10的架构示意图。终端400通过网络300连接服务器200,网络300可以是广域网或者局域网,又或者是二者的组合。终端400(运行有客户端,例如即时通信客户端、直播客户端、短视频客户端、社交客户端等)可以被用来获取针对用户的信息推荐请求,例如,当目标对象打开终端上运行的新闻客户端后,终端自动获取针对目标对象的新闻推荐请求。

[0106] 在一些实施例中,终端获取信息推荐请求后,调用服务器200的信息推荐接口(可

以提供为云服务的形式,即信息推荐服务),服务器200基于信息推荐请求,获取目标对象的多个参考特征,其中,参考特征包括以下至少之一:目标对象的对象特征、待推荐信息的信息特征;从待推荐信息库中召回符合上述特征的待推荐信息作为候选信息进行粗排。

[0107] 在重排阶段,分别对召回的待推荐信息的多个参考特征进行编码处理,得到各参考特征的编码特征;基于各编码特征,确定目标对象针对待推荐信息在至少两个推荐维度的第一推荐分数;对各编码特征在各推荐维度进行映射处理,得到相应的映射特征;将各推荐维度的第一推荐分数及相应的映射特征进行融合处理,得到对应的融合特征,并基于融合特征对待推荐信息的推荐分数进行预测,得到目标对象针对待推荐信息的第二推荐分数;基于第二推荐分数对召回的待推荐信息进行重排,并选择排名靠前的top N的待推荐信息推送至终端400显示。

[0108] 需要说明的是,本申请实施例中涉及的目标对象是信息推荐系统推荐信息的接收者,如目标对象打开新闻客户端时,目标对象即为新闻推荐系统推荐新闻的接收者,本申请实施例所涉及的目标对象的对象特征均是在征得目标对象同意的情况下获取的。

[0109] 在一些实施例中,本申请实施例提供的信息推荐方法还可应用于与车联网服务(如加油、导航、停车、维修等)相关的信息推荐场景中,如在对车载终端进行信息推荐时,对车载终端的目标对象执行本申请实施例提供的信息推荐方法,确定目标对象针对待推荐信息的最终推荐分数,并基于最终推荐分数执行待推荐信息对应目标对象的推荐;例如,对最终推荐分数低于分数阈值的待推荐信息应用相应的屏蔽模式,将最终推荐分数超过分数阈值的待推荐信息推荐给车载终端,从而避免质量低下的信息的广泛传播,间接提高整体信息质量,提高了用户体验。

[0110] 下面说明本申请实施例提供的用于信息推荐的电子设备的结构,参见图3,图3是本申请实施例提供的用于信息推荐的电子设备500的结构示意图,以电子设备500是服务器为例说明,图3所示的用于信息推荐的电子设备500包括:至少一个处理器510、存储器550、至少一个网络接口520和用户接口530。电子设备500中的各个组件通过总线系统540耦合在一起。可理解,总线系统540用于实现这些组件之间的连接通信。总线系统540除包括数据总线之外,还包括电源总线、控制总线和状态信号总线。但是为了清楚说明起见,在图3中将各种总线都标为总线系统540。

[0111] 处理器510可以是一种集成电路芯片,具有信号的处理能力,例如通用处理器、数字信号处理器(DSP, Digital Signal Processor),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,其中,通用处理器可以是微处理器或者任何常规的处理器等。

[0112] 存储器550包括易失性存储器或非易失性存储器,也可包括易失性和非易失性存储器两者。其中,非易失性存储器可以是只读存储器(ROM, Read Only Memory),易失性存储器可以是随机存取存储器(RAM, Random Access Memory)。本申请实施例描述的存储器550旨在包括任意适合类型的存储器。存储器550可选地包括在物理位置上远离处理器510的一个或多个存储设备。

[0113] 在一些实施例中,存储器550能够存储数据以支持各种操作,这些数据的示例包括程序、模块和数据结构或者其子集或超集,下面示例性说明。

[0114] 操作系统551,包括用于处理各种基本系统服务和执行硬件相关任务的系统程序,

例如框架层、核心库层、驱动层等,用于实现各种基础业务以及处理基于硬件的任务;

[0115] 网络通信模块552,用于经由一个或多个(有线或无线)网络接口520到达其他计算设备,示例性的网络接口520包括:蓝牙、无线相容性认证(WiFi)、和通用串行总线(USB, Universal Serial Bus)等;

[0116] 呈现模块553,用于经由一个或多个与用户接口530相关联的输出装置531(例如,显示屏、扬声器等)使得能够呈现信息(例如,用于操作外围设备和显示内容和信息的用户接口);

[0117] 输入处理模块554,用于对一个或多个来自一个或多个输入装置532之一的一个或多个用户输入或互动进行检测以及翻译所检测的输入或互动。

[0118] 在一些实施例中,本申请实施例提供的信息推荐装置可以采用软件方式实现,例如,可以是上文所述的服务器中信息推荐服务,还可以是上文所述的终端中信息推荐插件。当然,不局限于此,本申请实施例提供的信息推荐装置可以提供为各种软件实施例,包括应用程序、软件、软件模块、脚本或代码在内的各种形式。

[0119] 在一些实施例中,本申请实施例提供的信息推荐装置可以采用软件方式实现,图3示出了存储在存储器550中的信息推荐装置555,其可以是程序和插件等形式的软件,包括以下软件模块:特征编码模块5551、第一预测模块5552、特征映射模块5553、第二预测模块5554和信息推荐模块5555,这些模块是逻辑上的,因此根据所实现的功能可以进行任意的组合或进一步拆分。将在下文中说明每个模块的功能。

[0120] 在另一些实施例中,本申请实施例提供的信息推荐装置可以采用硬件方式实现,作为示例,本申请实施例提供的信息推荐装置可以是采用硬件译码处理器形式的处理器,其被编程以执行本申请实施例提供的信息推荐方法,例如,硬件译码处理器形式的处理器可以采用一个或多个应用专用集成电路(ASIC, Application Specific Integrated Circuit)、DSP、可编程逻辑器件(PLD, Programmable Logic Device)、复杂可编程逻辑器件(CPLD, Complex Programmable Logic Device)、现场可编程门阵列(FPGA, Field-Programmable Gate Array)或其他电子元件。

[0121] 下面将结合附图对本申请实施例提供的信息推荐方法进行说明,下述信息推荐方法的执行主体可以为服务器,具体可以是服务器通过运行上文的各种计算机程序来实现的;当然,根据对下文的理解,不难看出也可以由终端和服务器协同实施本申请实施例提供的信息推荐方法。

[0122] 参见图4A,图4A是本申请实施例提供的信息推荐方法的流程示意图,将结合图4A示出的步骤进行说明。

[0123] 在步骤101中,服务器分别对目标对象的多个参考特征进行编码处理,得到各参考特征的编码特征。

[0124] 其中,参考特征包括以下至少之一:目标对象的对象特征、待推荐信息的信息特征,对象特征是目标对象的基础属性特征(如年龄、性别、职业、受教育程度、消费水平等)、画像特征(如兴趣爱好、浏览、点击、收藏、购买等行为数据等)、以及推荐所处的上下文特征(环境特征,如推荐时间、推荐场景等),信息特征是待推荐信息的信息标签、信息类别、信息来源、信息内容与对象特征相关的交互特征,交互特征是指信息特征与对象特征的交集,如统计待推荐信息在不同年龄、性别用户的消费情况,通过具体的用户年龄、性别,得到该用

户年龄、性别下的用户对待推荐信息的消费特征,该消费特征即为待推荐信息与用户之间的交互特征。

[0125] 在实际实施时,可从中选择任意一个特征或多个特征进行自由组合作为参考特征,也即参考特征的规模维度可设置,如将目标对象的所有或部分对象特征作为参考特征,或将待推荐信息的所有或部分信息特征作为参考特征,或将目标对象的所有对象特征和待推荐信息的所有信息特征作为参考特征,或将目标对象的部分对象特征和待推荐信息的所有信息特征作为参考特征,将目标对象的部分对象特征和待推荐信息的部分信息特征作为参考特征,等等。

[0126] 在一些实施例中,参考特征包括连续型特征及离散型特征,步骤101可通过如下方式实现:对连续型特征进行离散化处理,得到连续型特征的离散特征;对连续型特征的离散特征进行编码处理,得到连续型特征的编码特征;对离散型特征进行编码处理,得到离散型特征的编码特征;将连续型特征的编码特征及离散型特征的编码特征确定为各参考特征的编码特征。

[0127] 这里,不同的参考特征,编码过程是不同的,参见图5,图5为本申请实施例提供的信息推荐示意图,假设待推荐信息为视频,参考特征中的连续性特征(特征数值具有连续性)可为视频时长、发布实效等,离散型特征(数值具有离散性)可为视频标识、用户账号等级、用户性别等,需要对不同的特征进行针对性的编码处理。例如,对视频时长、发布实效等连续性特征进行离散化,或对连线型特征进行归一化或标准化处理,得到连续性特征的离散特征(即离散数值,又称稀疏特征),再对连续性特征的离散特征进行编码处理,如通过embedding向量转化处理,得到连续性特征的编码特征(嵌入向量);对于视频标识、用户账号等级、用户性别等离散型特征直接进行编码处理,如通过embedding向量转化处理,得到离散性特征的编码特征(嵌入向量,又称稠密特征),从而通过针对性的编码处理,获得准确的编码特征,以便后续进行处理。

[0128] 在步骤102中,基于各编码特征,确定目标对象针对待推荐信息在至少两个推荐维度的第一推荐分数。

[0129] 在一些实施例中,参见图4B,图4B为本申请实施例提供的第一推荐分数的确定示意图,步骤102可通过图4B示出的步骤1021至步骤1024实现:

[0130] 在步骤1021中,对各编码特征进行特征交叉,得到对应的交叉特征。

[0131] 在一些实施例中,服务器可通过如下方式对各编码特征进行特征交叉,得到对应的交叉特征:对多个参考特征进行一阶特征处理,得到对应的一阶特征;对多个编码特征中任意两个编码特征进行二阶交叉处理,得到任意两个编码特征的二阶交叉特征;将得到的一阶特征及二阶交叉特征进行融合处理,得到对应的交叉特征。

[0132] 如图5所示,分别多个参考特征(即原始特征)进行一阶计算,如基于每个参考特征的权重,将待推荐信息的多个参考特征进行加权求和和处理,得到对应的一阶特征;对任意两个编码特征进行二阶特征交叉处理,得到任意两个编码特征的二阶交叉特征,如参考特征的编码特征 $x_{iu} = [x_{iu1}, x_{iu2}, \dots, x_{iun}]$, 其中, m 为参考特征的维度,则任意两个编码特征的二阶交叉特征 $p_{i,j} = \langle x_{iua}, x_{iub} \rangle$, $a = 1, \dots, m$, $b = 1, \dots, m$, 对所有任意两个编码特征

的二阶交叉特征 $\mathbf{p} = \sum (p_{i,j})$ 以及参考特征的一阶特征进行拼接处理,得到低阶具有记忆性的交叉特征。

[0133] 在一些实施例中,还可根据实际情况获取任意指定阶交叉特征,如对多个参考特征分别对应的编码特征中任意 i ($2 \leq i \leq m$, m 为参考特征的个数) 个编码特征进行 i 阶交叉处理,得到任意 i 个编码特征的 i 阶交叉特征,将一阶特征及 i 阶交叉特征进行拼接处理,得到对应的交叉特征,以便后续第一推荐评分的预测。

[0134] 在一些实施例中,为了提高处理速度,还可将一阶特征及 i 阶交叉特征进行拼接处理得到的拼接特征进行矩阵分解处理,如当 $i=2$ 时,对所有任意两个编码特征的二阶交叉特征以及参考特征的一阶特征进行拼接得到的拼接特征进行矩阵分解处理,得到分解特征,通过激活函数对分解特征进行非线性映射处理,得到对应的交叉特征。

[0135] 通过上述方式,通过对各编码特征进行特征交叉处理,捕捉不同编码特征之间的交叉信息,增强编码特征表征能力,避免遗漏特征边界,以便后续基于准确的交叉特征进行后续的预测处理。

[0136] 在步骤1022中,基于各编码特征,对待推荐信息在至少两个推荐维度的拟合性进行预测,得到对应各推荐维度的拟合特征。

[0137] 其中,拟合特征为用于表征所有推荐维度中任意两个推荐维度之间的拟合自由度的特征,通常为高阶特征。

[0138] 在一些实施例中,参见图4C,图4C为本申请实施例提供的拟合特征的确定示意图,步骤1022可通过图4C示出的步骤10221至步骤10223实现:步骤10221中,通过多门控混合专家网络中对应各推荐维度的专家网络,对各编码特征进行全连接处理,得到对应的第一隐层特征,并对第一隐层特征进行映射处理,得到对应各专家网络的映射特征;步骤10222中,通过多门控混合专家网络中对应各推荐维度的门控网络,对各编码特征进行全连接处理,得到对应的第二隐层特征,并对第二隐层特征进行映射处理,得到对应各专家网络的权重特征;步骤10223中,基于对应各专家网络的权重特征,对对应各专家网络的映射特征进行加权求和处理,得到对应各推荐维度的拟合特征。

[0139] 其中,多门控混合专家网络由多个专家网络和多个门控网络构成,专家网络用于提取不同的特征,可为DNN网络结构,门控网络用于分配每个专家网络的权重,每个门控网络相当于一个分类器,每个推荐维度的门控网络会根据当前输入的编码特征,判别它由哪些专家网络来拟合会更好,故预估出每个专家网络的权重。专家网络和门控网络的数量可根据实际情况设置,如数量可与推荐维度的数量相一致,即每个推荐维度对应一个专家网络和一个门控网络,如推荐维度为点击、时长和互动三个维度时,“点击”这一推荐维度对应专家网络1和门控网络1,“时长”这一推荐维度对应专家网络2和门控网络2,“互动”这一推荐维度对应专家网络3和门控网络3。

[0140] 如图5所示,将多个参考特征对应的编码特征分别输入至各推荐维度对应的专家网络中,首先通过对应各推荐维度的专家网络,对各编码特征进行全连接处理,得到对应的第一隐层特征,并通过激活函数对第一隐层特征进行线性或非线性映射处理,得到对应各专家网络的映射特征;然后将各个编码特征以及各个专家网络输出的映射特征输入至相应的门控网络中,通过门控网络各个编码特征进行MLP处理,如全连接处理,得到对应的第二

隐层特征,然后通过激活函数对第二隐层特征进行线性或非线性映射处理,得到对应各专家网络的权重特征,并以此权重特征,对对应各专家网络的映射特征进行加权求和处理,得到对应各推荐维度的拟合特征,作为各推荐维度对应的门控网络的输出,并将各推荐维度对应的门控网络的输出作为整个多门控混合专家网络的输出。

[0141] 在步骤1023中,分别将交叉特征及对应各推荐维度的拟合特征进行拼接处理,得到对应各推荐维度的拼接特征。

[0142] 这里,将低阶具有记忆性的交叉特征与对应各推荐维度的的高阶特征进行拼接后进行后续的分数的预测。

[0143] 如图5所示,在获得各推荐维度对应的门控网络输出的拟合特征后,分别将交叉特征及对应各推荐维度的拟合特征进行拼接处理,得到对应各推荐维度的拼接特征。仍以点击、时长和互动这三个推荐维度为例,将上述得到的交叉特征与“点击”这一推荐维度的拟合特征进行拼接,得到“点击”这一推荐维度的拼接特征;将上述得到的交叉特征与“时长”这一推荐维度的拟合特征进行拼接,得到“时长”这一推荐维度的拼接特征;将上述得到的交叉特征与“互动”这一推荐维度的拟合特征进行拼接,得到“互动”这一推荐维度的拼接特征。

[0144] 在步骤1024中,基于拼接特征,针对待推荐信息在至少两个推荐维度的推荐分数进行预测,得到对应各推荐维度的第一推荐分数。

[0145] 在一些实施例中,针对各推荐维度,对相应的拼接特征进行映射处理,得到对应的映射特征,通过激活函数对映射特征进行偏置处理,得到对应各推荐维度的第一推荐分数。在另一些实施例中,针对各推荐维度,将相应的拼接特征输入至分数预测模型中,通过分数预测模型对拼接特征进行投影处理,得到对应的投影特征,如通过逻辑回归函数对拼接特征进行线性逻辑回归处理,这里的线性逻辑回归处理可以是线性加和处理,也可以是对线性加和结果代入逻辑回归函数,得到逻辑回归特征作为投影特征,然后经过激活函数对投影特征进行推荐分数预测处理,得到表征推荐分数高低的第一推荐分数。

[0146] 仍以上述示例为例,在分别得到对应点击、时长和互动这三个推荐维度的拼接特征后,分别预测得到“点击”这一推荐维度的第一推荐分数、“时长”这一推荐维度的第一推荐分数,以及“互动”这一推荐维度的第一推荐分数。

[0147] 在步骤103中,对各编码特征在各推荐维度进行映射处理,得到相应的映射特征。

[0148] 其中,映射特征用于表征在相应的推荐维度所述第一推荐分数的融合权重。

[0149] 在一些实施例中,步骤103可通过如下方式实现:对各推荐维度的第一推荐分数进行横向拼接处理,得到对应的平铺向量;对各编码特征进行全连接处理,得到对应的隐层特征;对隐层特征进行映射处理,得到与平铺向量维度相同的映射特征。

[0150] 这里,在实际实施时,在得到对应每个推荐维度的第一推荐分数之后,对各第一推荐分数的向量表示进行横向拼接处理,得到第一推荐分数对应的平铺向量,记为

$x_{in} = [x_{i1}, x_{i2}, \dots, x_{in}]$, 其中, x_{in} 表示第n个推荐维度的第一推荐分数的向量表示,n

表示推荐维度的个数;然后,将各编码特征降维至与平铺向量维度规模相同的映射特征,如将各编码特征通过全连接处理,得到对应的隐层特征,并通过激活函数对隐层特征进行非线性映射处理,得到与平铺向量维度规模相同的映射特征,以便于后续映射特征与各推荐

维度的第一推荐分数的关联计算。

[0151] 在步骤104中,将各推荐维度的第一推荐分数及相应的映射特征进行融合处理,得到对应的融合特征,并基于融合特征对待推荐信息的推荐分数进行预测,得到目标对象针对待推荐信息的第二推荐分数。

[0152] 在一些实施例中,步骤104中将各推荐维度的第一推荐分数及相应的映射特征进行融合处理,得到对应的融合特征,可通过如下方式实现:获取由各推荐维度的第一推荐分数构成的分数矩阵,以及各推荐维度对应的映射特征构成的映射矩阵;将分数矩阵与映射矩阵进行元素积计算,得到对应的融合特征。

[0153] 其中,分数矩阵即为上述对各第一推荐分数的向量表示进行横向拼接处理得到的平铺向量 x_{in} ,映射矩阵为与平铺向量维度规模相同的映射特征,记为 w_{iu} ,将分数矩阵与映射矩阵进行元素积计算得到的融合特征记为: $w_{i,um} = w_{iu} \odot x_{in}$,以衡量目标对象在不同推荐维度下的倾向大小。

[0154] 在一些实施例中,步骤104中基于融合特征对待推荐信息的推荐分数进行预测,得到目标对象针对所述待推荐信息的第二推荐分数,可通过如下方式实现:对融合特征进行映射处理,得到对应的映射特征;基于映射特征对待推荐信息的推荐分数进行预测,得到目标对象针对待推荐信息的第二推荐分数。

[0155] 在进行推荐分数预测时,对融合特征进行映射处理,如将融合特征通过逻辑回归函数做线性投影,得到对应的投影值(映射特征),然后将得到的投影值经过激活函数预测得到目标对象针对待推荐信息的第二推荐分数。

[0156] 在步骤105中,基于第二推荐分数,执行待推荐信息对应目标对象的推荐。

[0157] 其中,第二推荐分数是综合多个推荐维度的最终得分,用于表征目标对象针对待推荐信息的整体评价,当第二推荐分数超过分数阈值时,将待推荐信息推荐给目标对象。

[0158] 在一些实施例中,本申请实施例提供的信息推荐方法可适用于在推荐系统的召回阶段,在获取召回的每个候选信息的第二推荐分数后,按照第二推荐分数由高到低的顺序,对召回的候选信息进行重排,并选择排名靠前的top N的候选信息推送至终端显示。

[0159] 在一些实施例中,还可根据第二推荐分数可确定目标对象针对待推荐信息的评价等级,进而根据评价等级进行不同的推荐操作。例如,假设评价等级包括等级依次增高(用户越来越感兴趣)的第一等级、第二等级和第三等级时,当针对待推荐信息的评价等级为第一等级时,在推荐系统的排序阶段,对待推荐信息进行降权推荐,以减少推荐次数或推荐频率,如在未采取降权排序之前,在一周内可能会向100个人推荐该信息,在采取降权排序之后,在一周内可能只向20个人推荐该信息,另外,降权的幅度与待推荐信息的最终得分呈负相关关系,即待推荐信息的最终得分越低,降权幅度越大,降权排序后在一定时间内针对该信息的推荐次数或推荐频率就越低;在推荐系统的召回阶段,将包含待推荐信息的召回结果中对待推荐信息进行暂时过滤或永久过滤,然后基于过滤后的信息进行后续的排序再推荐,以避免将用户不感兴趣的信息推荐给目标对象或与目标对象相似的其他用户。

[0160] 当针对待推荐信息的评价等级为第二等级时,对待推荐信息进行自由推荐,自由推荐即不对待推荐信息进行偏向性推荐,既不加权推荐,也不降权推荐,使其基于用户需求和信息自身质量进行推荐。当针对待推荐信息的评价等级为第三等级时,对待推荐信息进

行加权推荐,从而使目标对象感兴趣的待推荐信息可以被推荐给更多与目标对象相似的其他用户,增加待推荐信息的曝光率和点击率。

[0161] 在一些实施例中,上述信息推荐方法是通过调用分数预测模型实现的,如图5所示,分数预测模型包括:特征编码层、第一推荐分数预测层、特征映射层和第二推荐分数预测层;其中,第一推荐分数预测层包括第一特征提取层、第二特征提取层、特征拼接层和子分数预测层;第二推荐分数预测层包括特征融合层和总分数预测层。

[0162] 在一些实施例中,参见图6A,图6A为本申请实施例提供的模型训练方法流程示意图,可通过如下方式对分数预测模型进行训练:在步骤201中,服务器通过特征编码层,分别对训练样本的多个参考特征进行编码处理,得到对应各参考特征的编码特征,训练样本携带对象样本针对信息样本在至少两个推荐维度的第一标签,以及对象样本针对信息样本的第二标签;在步骤202中,通过第一推荐分数预测层,基于各编码特征,确定对象样本针对信息样本在至少两个推荐维度的第一预测结果;在步骤203中,通过特征映射层,对各编码特征在各推荐维度进行映射处理,得到相应的映射特征,其中,映射特征用于表征第一推荐分数在相应的推荐维度的融合权重;在步骤204中,通过第二推荐分数预测层,将各推荐维度的第一推荐分数及相应的映射特征进行融合处理,得到对应的融合特征,并基于融合特征,对信息样本的推荐分数进行预测,得到对象样本针对所述信息样本的第二预测结果;在步骤205中,基于各推荐维度的第一预测结果与相应的第一标签,及第二预测结果与所述第二标签,更新分数预测模型的模型参数。

[0163] 在实际实施时,将训练样本输入至分数预测模型中,首先,通过特征编码层将训练样本的多个参考特征进行编码处理,如进行嵌入压缩处理,以将多个参考特征的稀疏特征转换为稠密特征。其次,通过第一推荐分数预测层中的第一特征提取层,对训练样本的各编码特征进行特征交叉,得到对应的交叉特征;通过第二特征提取层,基于各编码特征对待推荐信息在至少两个推荐维度的拟合性进行预测,得到对应各推荐维度的拟合特征;通过特征拼接层,分别将交叉特征及对应各推荐维度的拟合特征进行拼接处理,得到对应各推荐维度的拼接特征;通过子分数预测层,基于拼接特征针对待推荐信息在至少两个推荐维度的推荐分数进行预测,得到对象样本针对信息样本在至少两个推荐维度的第一预测结果。再次,通过特征映射层,对训练样本的各编码特征在各推荐维度进行映射处理,得到相应的映射特征。最后,通过第二分数预测层中的特征融合层,将对象样本针对信息样本在至少两个推荐维度的第一预测结果及相应的映射特征进行融合处理,得到对应的融合特征;通过总分数预测层,基于融合特征对待推荐信息的推荐分数进行预测,得到目标对象针对待推荐信息的第二预测结果。

[0164] 在一些实施例中,参见图6B,图6B为本申请实施例提供的模型训练方法流程示意图,步骤205可通过图6B示出的步骤2051至步骤2054实现:在步骤2051中,针对各推荐维度,基于第一预测结果及相应的第一标签,构造对应第一推荐分数预测层的第一损失函数;在步骤2052中,基于第二预测结果及第二标签,构造对应第二推荐分数预测层的第二损失函数;在步骤2053中,将第二损失函数及第一损失函数进行加权求和,得到分数预测模型的第三损失函数;在步骤2054中,基于第三损失函数更新分数预测模型的模型参数。

[0165] 在一些实施例中,上述步骤2051可通过如下方式实现:基于各推荐维度对应的第一预测结果及相应的第一标签,构造对应各推荐维度的子损失函数;确定各推荐维度对应

的推荐权重,基于各推荐权重对对应各推荐维度的子损失函数进行加权求和,得到对应第一推荐分数预测层的第一损失函数。

[0166] 这里,对于每个推荐维度,在得到对应的第一预测结果后,可基于第一预测结果与训练样本携带的对象样本针对信息样本在相应推荐维度的第一标签,构建对应的子损失函数,并将各个推荐维度的子损失函数相加得到第一推荐分数预测层的第一损失函数

$L_1 = \sum_{j=1}^n loss_j$, 其中, n 表示推荐维度的个数, $loss_j$ 表示第 j 个推荐维度对应的子损失函数, $1 \leq j \leq n$ 。

[0167] 在得到综合每个推荐维度最终的第二预测结果后,可基于第二预测结果与训练样本携带的对象样本针对信息样本的第二标签,构建第二推荐分数预测层的第二损失函数 $L(\theta)$ 表示为:

[0168]
$$L(\theta) = \sum_{i=1}^n -a_i \left[y_i \log p(w_{i,um} | \theta) + (1 - y_i) \log (1 - p(w_{i,um} | \theta)) \right];$$

[0169] 其中, $p(w_{i,um} | \theta) = \sigma(f(w_{i,um} | \theta))$, $\sigma(\cdot)$ 为sigmoid函数, $f(w_{i,um} | \theta)$ 为第二预测结果,将第二预测结果转换为预估概率 $p(w_{i,um} | \theta)$, n 为训练样本的总数, θ 为模型参数, y_i 为第二标签, a_i 为根据推荐维度设置的不同权重,如一个训练样本有点击和互动两个推荐维度, a_i 可设置为2,比单纯仅有点击这一个推荐维度的训练样本的权重(小于2)大,模型更倾向学习带互动的训练样本。

[0170] 分数预测模型整体的第三损失函数为第一损失函数(即多个推荐维度中每个推荐维度单独的子损失函数之和)和第二损失函数相加表示: $Loss = L_1 + L(\theta) = \sum_{j=1}^n loss_j + L(\theta)$ 。

[0171] 在构建第三损失函数后,根据第三损失函数的值判断第三损失函数的值是否超出预设阈值,当第三损失函数的值超出预设阈值时,基于第三损失函数确定分数预测模型的误差信号,将误差信息在分数预测模型中反向传播,并在传播的过程中更新各个层的模型参数。

[0172] 这里,对反向传播进行说明,将训练样本的参考特征输入到神经网络模型的输入层,经过隐藏层,最后达到输出层并输出结果,这是神经网络模型的前向传播过程,由于神经网络模型的输出结果与实际结果有误差,则计算输出结果与实际值之间的误差,并将该误差从输出层向隐藏层反向传播,直至传播到输入层,在反向传播的过程中,根据误差调整模型参数的值;不断迭代上述过程,直至收敛,其中,分数预测模型属于神经网络模型。

[0173] 下面,将说明本申请实施例在一个实际的应用场景中的示例性应用。本申请实施例提供的信息推荐方法可应用于一切使用多目标排序模型的推荐系统上,如可应用于客户端推荐、浏览器信息流场景、新闻、快报推荐等信息流产品,还可以应用于如电商领域、广告推荐场景等其他推荐场景。接下来以点击、时长、互动这三个推荐维度的多目标分数融合为例,对本申请实施例提供的信息推荐方法进行说明。

[0174] 参见图7,图7为本申请实施例提供的信息推荐示意图,通过多目标排序模型进行信息推荐,该模型包括:稀疏特征层、特征提取层、子得分预测层、特征映射层、融合部分,接下来将结合图7对分数预测模型的训练和应用进行说明。

[0175] 1、稀疏特征层

[0176] 在选择训练样本的用户侧特征(即上述的参考特征)时,可从对象样本的用户特征(即上述的对象特征)、待推荐信息的信息特征中选择,其中,用户特征是目标对象的基础属性特征(如年龄、性别、职业、受教育程度、消费水平等)、画像特征(如兴趣爱好、浏览、点击、收藏、购买等行为数据等)、以及推荐所处的上下文特征(环境特征,如推荐时间、推荐场景等),信息特征是信息样本的信息标签、信息类别、信息来源、信息内容与用户特征的交互特征,交互特征是指信息特征与用户特征的交集,如统计待推荐信息在不同年龄、性别用户的消费情况,通过具体的用户年龄、性别,得到该用户年龄、性别下的用户对待推荐信息的消费特征,该消费特征即为待推荐信息与用户之间的交互特征。

[0177] 当用户侧特征中存在连续型特征时,需先对连续型特征进行离散化处理,或对连续型特征进行归一化或标准化处理得到离散特征;通常情况下,离散特征为稀疏特征,需通过稀疏特征层对离散特征进行编码处理,如通过embedding向量转化处理,得到对应的编码特征(又称稠密特征);将得到的编码特征及用户侧本身即为稠密特征的用户侧特征进行拼接,得到用户侧特征向量(即编码特征)为 $x_{iu} = [x_{iu1}, x_{iu2}, \dots, x_{iun}]$,其中,m为用户侧特征的个数。

[0178] 2、特征提取层

[0179] 特征提取层包括交叉特征提取层和拟合特征提取层,其中,交叉特征提取层可为因子分解机(FM, Factorization Machine)模型,拟合特征提取层可为MMoE模型,通过交叉特征提取层,对用户侧特征的各编码特征进行二阶特征交叉,得到对应的二阶交叉特征,将各编码特征的一阶特征与二阶交叉特征进行拼接,得到低阶具有记忆性的交叉特征;MMoE模型由多个专家网络和多个门控网络构成,专家网络用于提取不同的特征,可为DNN网络结构,门控网络用于分配每个专家网络的权重,每个门控网络相当于一个分类器,每个推荐维度的门控网络会根据当前输入的编码特征,判别它由哪些专家网络来拟合会更好,故预估出每个专家网络的权重。最后,将低阶具有记忆性的交叉特征与MMoE模型输出的对应各个目标的高阶特征进行拼接后,输入至子分数预测层中进行分数预测。

[0180] 3、子分数预测层

[0181] 其中,子分数预测层包括对点击、时长、互动进行分数预测的三个模型,三个模型相互独立,将交叉特征与MMoE模型输出的对应各个目标的输出进行拼接后输入至相应的模型中进行分数预测,得到相应的得分(即上述的第一推荐分数)。

[0182] 4、特征映射层

[0183] 在得到对应每个目标的得分之后,对各得分的向量表示进行横向拼接处理,得到多目标得分向量,记为 $x_{im} = [x_{i1}, x_{i2}, \dots, x_{im}]$,其中, x_{in} 表示第n个目标的得分的向量表示,n表示目标的个数;然后,通过特征映射层将用户侧特征的编码特征降维至与多目标得分向量维度规模相同的矩阵,记为 w_{iu} ,其中,特征映射层可为MLP网络,如DNN网络;然后将多目

标得分向量与降维后用户侧特征的矩阵进行元素积计算得到的融合特征记为：

$$w_{i,um} = w_{iu} \odot x_{im}, \text{以衡量目标对象在不同目标下的倾向大小。}$$

[0184] 需要说明的是,特征映射层实质为用户的轻量网络,输入到特征映射层的特征可以来源于稀疏特征层输出的编码特征,即输入到特征映射层的特征可以是稀疏特征层输出的编码特征的部分或全部,也可以是其他新的特征,如获取与输入到稀疏特征层的用户侧特征不同的用户侧特征,新获取的用户侧特征中甚至可以包括信息特征,并对新获取的用户侧特征进行编码处理后输入至特征映射层。

[0185] 通过上述方式,用户个性化特征的引入可以根据不同用户给出所有目标分数的最优融合方式,相当于自动根据用户在不同目标上的倾向给出最终得分,在业务表现上达到相对更优的效果。

[0186] 5、融合部分

[0187] 融合部分的作用在于预测目标对象针对推荐信息的最终得分,在实际实施时,将上述融合特征 $w_{i,um} = w_{iu} \odot x_{im}$ 通过DNN预测得到目标对象针对推荐信息的最终

得分(即上述的第二推荐分数): $\text{final}_{\text{score}} = f(w_{i,um} | \theta)$ 。

[0188] 6、损失函数

[0189] 这里,在得到对应每个目标的得分后,可基于得分与训练样本携带的对象样本针对信息样本在相应目标的标签,构建对应的子损失函数,并将各个目标的子损失函数相加

得到子分数预测层的损失函数 $L_1 = \sum_{j=1}^n \text{loss}_j$,其中,n表示目标的个数, loss_j 表示第j个目

标对应的子损失函数, $1 \leq j \leq n$ 。

[0190] 在得到最终得分后,可基于最终得分与训练样本携带的对象样本针对信息样本的标签,构建融合部分的损失函数 $L(\theta)$ 表示为:

$$[0191] \quad L(\theta) = \sum_{i=1}^n -a_i \left[y_i \log p(w_{i,um} | \theta) + (1 - y_i) \log (1 - p(w_{i,um} | \theta)) \right];$$

[0192] 其中, $p(w_{i,um} | \theta) = \sigma(f(w_{i,um} | \theta))$, $\sigma(\cdot)$ 为sigmoid函数, $f(w_{i,um} | \theta)$ 为最终得分,将最终得分转换为预估概率 $p(w_{i,um} | \theta)$, n为训练样本的总数, y_i 为标签, a_i 为目标设置的不同权重,如一个训练样本有点击和互动两个目标, a_i 可设置为2,比单纯仅有点击这一个目标的训练样本的权重(小于2)大,模型更倾向学习带互动的训练样本。

[0193] 多目标排序模型整体的损失函数为子分数预测层的损失函数和融合部分的损失函数之和,表示为: $\text{Loss} = L_1 + L(\theta) = \sum_{j=1}^n \text{loss}_j + L(\theta)$ 。

[0194] 在构建多目标排序模型整体的损失函数后,根据多目标排序模型整体的损失函数的值(如梯度值)判断是否超出预设阈值,当超出预设阈值时,基于多目标排序模型整体的

损失函数确定模型的误差信号,将误差信号在分数预测模型中反向传播,并在传播的过程中更新各个层的模型参数。

[0195] 可见,本申请实施例提供的多目标排序模型是一种端到端的模型,不需要考虑离线和在线两种场景下数据的分布影响,融合部分的损失函数和多目标其他损失函数联合训练,在线预测使用只需要加载一个模型,提高服务部署的便利性和稳定性。

[0196] 8、预测阶段

[0197] 以待推荐信息为文章为例,当用户请求时,将用户侧特征(包括用户特征、候选文章的信息特征、交叉特征和上下文特征等)输入到多目标排序模型中,得到当前用户对每篇候选文章的预估点击率,预估时长或时长概率,转换为得分构成多目标得分向量 x_{im} ,同时根据需要的用户侧特征构建特征映射层,经过MLP后输出 w_{iu} ,将 x_{im} 和 w_{iu} 进行点乘,得到点乘结果 $w_{i,um} = w_{iu} \odot x_{im}$,最后把 $w_{i,um}$ 输入到融合部分,得到用户对候选文章的最终得分 $final_{score} = f(w_{i,um} | \theta)$,根据最终得分从大到小的顺序,对所有候选文章进行排序,返回前面K篇文章作为结果呈现给用户。

[0198] 参见图8A-图8C,图8A-图8C为本申请实施例提供的信息推荐效果示意图,以将本申请实施例提供的信息推荐方法应用于看点图文推荐场景下为例,使用本申请实施例提供的信息推荐方法相较于,一般的公式融合和网格搜参方法,在点击率、阅读总时长和点赞人数这三个目标上的相对提升幅度,如点击率平均相对提升了1.16%,最高相对提升1.62%(图8A),时长平均相对提升了1.17%,最高相对提升1.38%(图8B)、点赞平均相对提升了2.76%,最高相对提升3.77%(图8C),其中,空跑期是用一般的公式融合和网格搜参方法,实验期为本申请实施例提供的信息推荐方法相对对比一般融合和网格搜参方法的提升效果。

[0199] 通过上述方式,本申请实施例提供一种端到端的基于用户个性化特征的多目标分数融合模型,旨在解决多目标模型在不同目标上的打分融合成一个得分进行排序的问题,本申请实施例提供的信息推荐方法在基于MMoE的多目标模型的基础上,引入MLP网络自适应学习多目标融合打分,即通过引入用户侧特征,可以自适应的学习每个用户到不同目标得分的个性化权重,综合用户侧特征得到最佳的融合得分,进而实现各个目标上的最优,不再依赖人工公式和搜参,省时省力。

[0200] 下面继续说明本申请实施例提供的信息推荐装置555的实施为软件模块的示例性结构,在一些实施例中,存储在图3中存储器550的信息推荐装置555中的软件模块可以包括:

[0201] 特征编码模块5551,用于分别对目标对象的多个参考特征进行编码处理,得到对应各所述参考特征的编码特征;

[0202] 第一预测模块5552,用于基于各所述编码特征,确定所述目标对象针对待推荐信息在至少两个推荐维度的第一推荐分数;

[0203] 特征映射模块5553,用于对各所述编码特征在各所述推荐维度进行映射处理,得到相应的映射特征,所述映射特征用于表征在相应的所述推荐维度所述第一推荐分数的融合权重;

[0204] 第二预测模块5554,用于将各所述推荐维度的第一推荐分数及相应的所述映射特

征进行融合处理,得到对应的融合特征,并基于所述融合特征对所述待推荐信息的推荐分数进行预测,得到所述目标对象针对所述待推荐信息的第二推荐分数;

[0205] 信息推荐模块5555,用于基于所述第二推荐分数,执行所述待推荐信息对应所述目标对象的推荐。

[0206] 在一些实施例中,所述参考特征包括连续型特征及离散型特征,所述特征编码模块5551,还用于对所述连续型特征进行离散化处理,得到所述连续型特征的离散特征;对所述连续型特征的离散特征进行编码处理,得到所述连续型特征的编码特征;对所述离散型特征进行编码处理,得到所述离散型特征的编码特征;将所述连续型特征的编码特征及所述离散型特征的编码特征确定为各所述参考特征的编码特征。

[0207] 在一些实施例中,所述第一预测模块5552,还用于对各所述编码特征进行特征交叉,得到对应的交叉特征;基于各所述编码特征,对所述待推荐信息在至少两个推荐维度的拟合性进行预测,得到对应各推荐维度的拟合特征;分别将所述交叉特征及对应各所述推荐维度的拟合特征进行拼接处理,得到对应各所述推荐维度的拼接特征;基于所述拼接特征,针对所述待推荐信息在至少两个推荐维度的推荐分数进行预测,得到对应各所述推荐维度的第一推荐分数。

[0208] 在一些实施例中,所述第一预测模块5552,还用于对所述多个参考特征进行一阶特征处理,得到对应的一阶特征;对多个所述编码特征中任意两个编码特征进行二阶交叉处理,得到所述任意两个编码特征的二阶交叉特征;将得到的所述一阶特征及所述二阶交叉特征进行融合处理,得到对应的交叉特征。

[0209] 在一些实施例中,所述第一预测模块5552,还用于通过多门控混合专家网络中对应各所述推荐维度的专家网络,对各所述编码特征进行全连接处理,得到对应的第一隐层特征,并对所述第一隐层特征进行映射处理,得到对应各所述专家网络的映射特征;通过所述多门控混合专家网络中对应各所述推荐维度的门控网络,对各所述编码特征进行全连接处理,得到对应的第二隐层特征,并对所述第二隐层特征进行映射处理,得到对应各所述专家网络的权重特征;基于对应各所述专家网络的权重特征,对对应各所述专家网络的映射特征进行加权求和处理,得到对应各所述推荐维度的拟合特征。

[0210] 在一些实施例中,所述特征映射模块5553,还用于对各所述推荐维度的第一推荐分数进行横向拼接处理,得到对应的平铺向量;对各所述编码特征进行全连接处理,得到对应的隐层特征;对所述隐层特征进行映射处理,得到与所述平铺向量维度相同的映射特征。

[0211] 在一些实施例中,所述第二预测模块5554,还用于获取由各所述推荐维度的第一推荐分数构成的分数矩阵,以及各所述推荐维度对应的映射特征构成的映射矩阵;将所述分数矩阵与所述映射矩阵进行元素积计算,得到对应的融合特征。

[0212] 在一些实施例中,所述第二预测模块5554,还用于对所述融合特征进行映射处理,得到对应的映射特征;基于所述映射特征对所述待推荐信息的推荐分数进行预测,得到所述目标对象针对所述待推荐信息的第二推荐分数。

[0213] 在一些实施例中,所述信息推荐方法是通过调用分数预测模型实现的,所述分数预测模型包括:特征编码层、第一推荐分数预测层、特征映射层和第二推荐分数预测层;所述装置还包括:模型训练模块,用于通过所述特征编码层,分别对训练样本的多个参考特征进行编码处理,得到对应各所述参考特征的编码特征,所述训练样本携带对象样本针对信

息样本在至少两个推荐维度的第一标签,以及所述对象样本针对所述信息样本的第二标签;通过所述第一推荐分数预测层,基于各所述编码特征,确定所述对象样本针对所述信息样本在至少两个推荐维度的第一预测结果;通过所述特征映射层,对各所述编码特征在各所述推荐维度进行映射处理,得到相应的映射特征,所述映射特征用于表征在相应的所述推荐维度所述第一推荐分数的融合权重;通过所述第二推荐分数预测层,将各所述推荐维度的第一推荐分数及相应的所述映射特征进行融合处理,得到对应的融合特征,并基于所述融合特征,对所述信息样本的推荐分数进行预测,得到所述对象样本针对所述信息样本的第二预测结果;基于各所述推荐维度的第一预测结果与相应的第一标签,及所述第二预测结果与所述第二标签,更新所述分数预测模型的模型参数。

[0214] 在一些实施例中,所述模型训练模块,还用于针对各所述推荐维度,基于所述第一预测结果及相应的第一标签,构造对应所述第一推荐分数预测层的第一损失函数;基于所述第二预测结果及所述第二标签,构造对应所述第二推荐分数预测层的第二损失函数;将所述第二损失函数及所述第一损失函数进行加权求和,得到所述分数预测模型的第三损失函数;基于所述第三损失函数更新所述分数预测模型的模型参数。

[0215] 在一些实施例中,所述模型训练模块,还用于基于各所述推荐维度对应的第一预测结果及相应的第一标签,构造对应各所述推荐维度的子损失函数;确定各所述推荐维度对应的推荐权重,基于各所述推荐权重对对应各所述推荐维度的子损失函数进行加权求和,得到对应所述第一推荐分数预测层的第一损失函数。

[0216] 本申请实施例提供了一种计算机程序产品或计算机程序,该计算机程序产品或计算机程序包括计算机指令,该计算机指令存储在计算机可读存储介质中。计算机设备的处理器从计算机可读存储介质读取该计算机指令,处理器执行该计算机指令,使得该计算机设备执行本申请实施例上述的信息推荐方法。

[0217] 本申请实施例提供一种存储有可执行指令的计算机可读存储介质,其中存储有可执行指令,当可执行指令被处理器执行时,将引起处理器执行本申请实施例提供的信息推荐方法,例如,如图4A示出的方法。

[0218] 在一些实施例中,计算机可读存储介质可以是FRAM、ROM、PROM、EPROM、EEPROM、闪存、磁表面存储器、光盘、或CD-ROM等存储器;也可以是包括上述存储器之一或任意组合的各种设备。

[0219] 在一些实施例中,可执行指令可以采用程序、软件、软件模块、脚本或代码的形式,按任意形式的编程语言(包括编译或解释语言,或者声明性或过程性语言)来编写,并且其可按任意形式部署,包括被部署为独立的程序或者被部署为模块、组件、子例程或者适合在计算环境中使用的其它单元。

[0220] 作为示例,可执行指令可以但不一定对应于文件系统中的文件,可以可被存储在保存其它程序或数据的文件的一部分,例如,存储在超文本标记语言(HTML, Hyper Text Markup Language)文档中的一个或多个脚本中,存储在专用于所讨论的程序的单个文件中,或者,存储在多个协同文件(例如,存储一个或多个模块、子程序或代码部分的文件)中。

[0221] 作为示例,可执行指令可被部署为在一个计算设备上执行,或者在位于一个地点的多个计算设备上执行,又或者,在分布在多个地点且通过通信网络互连的多个计算设备上执行。

[0222] 以上所述,仅为本申请的实施例而已,并非用于限定本申请的保护范围。凡在本申请的精神和范围之内所作的任何修改、等同替换和改进等,均包含在本申请的保护范围之内。

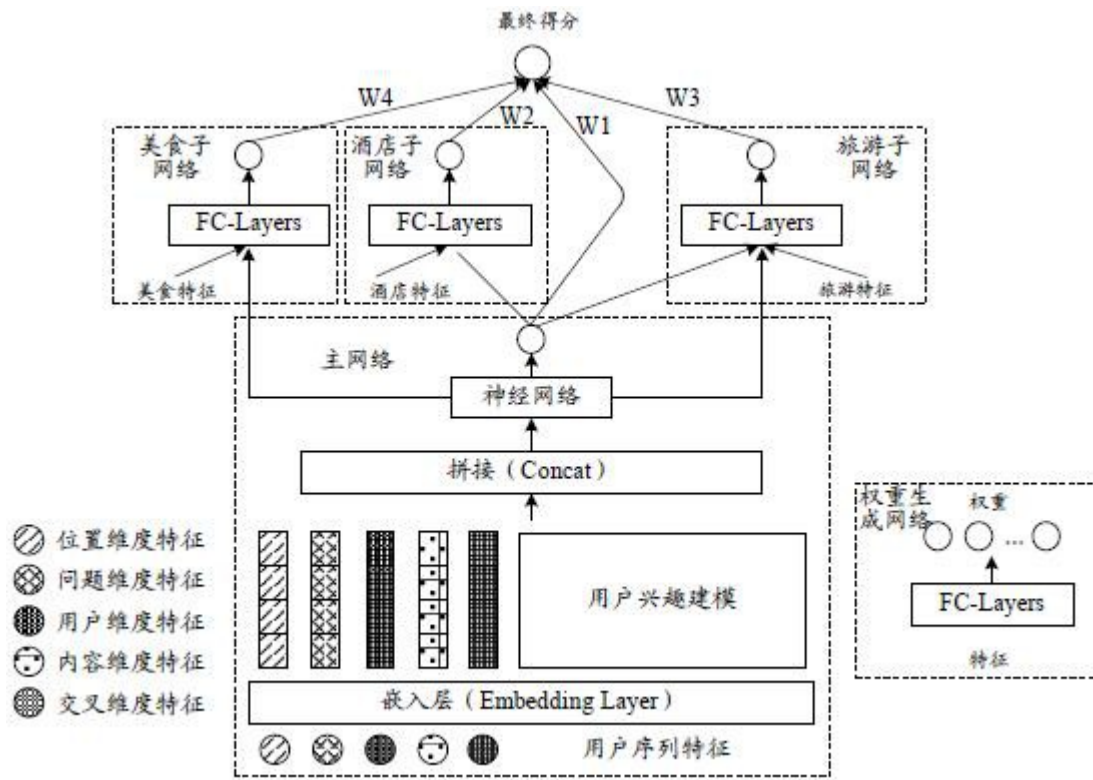


图1A

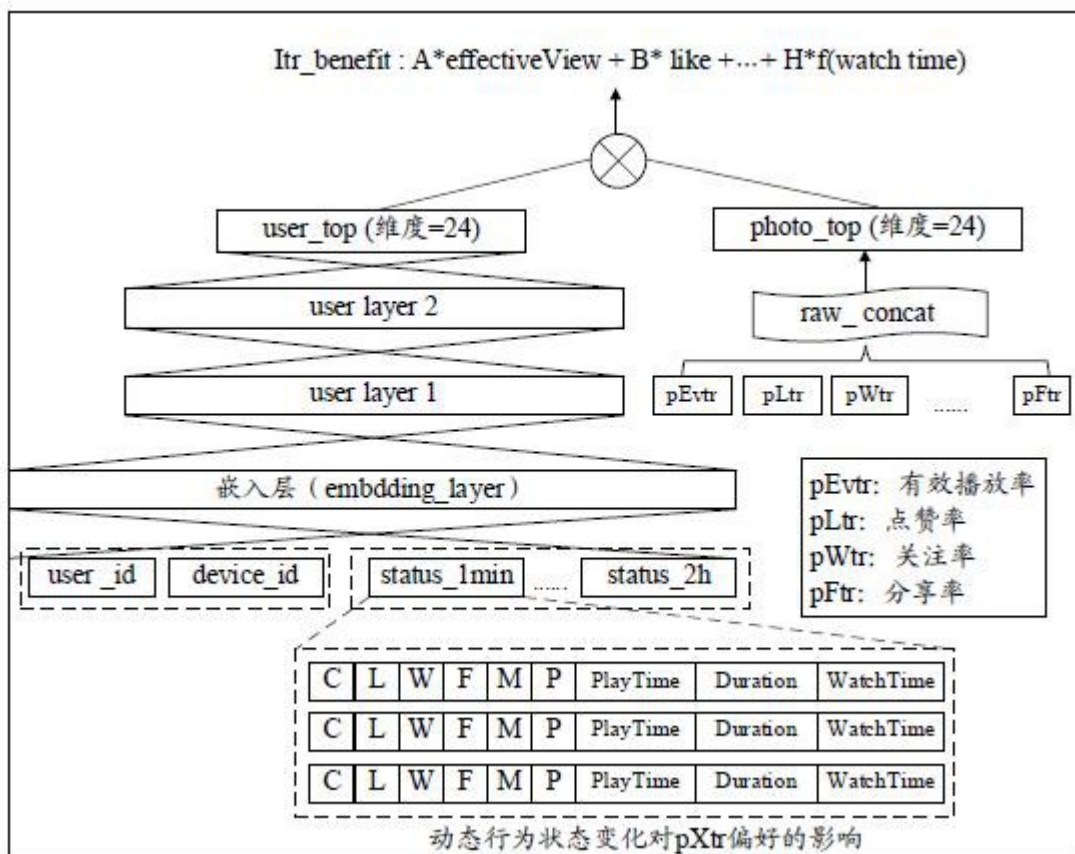


图1B

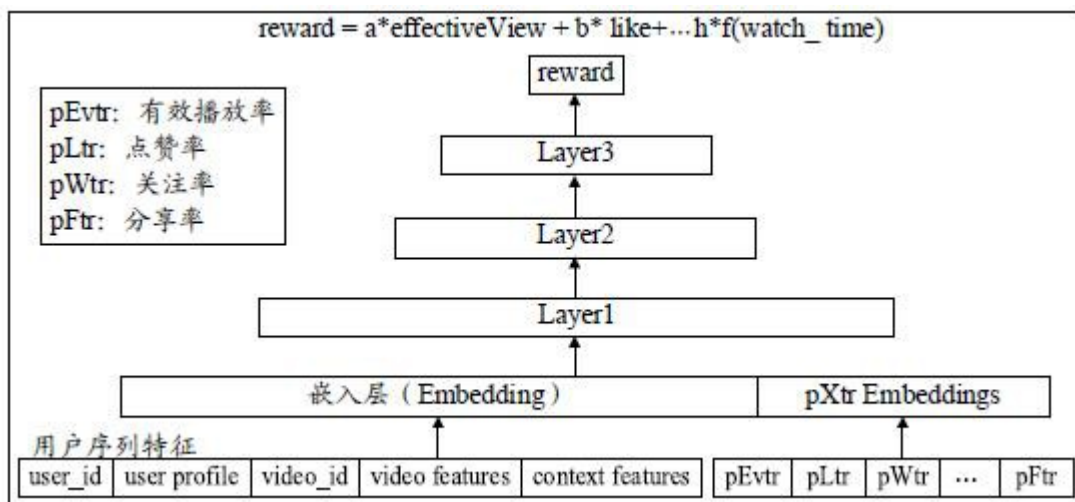


图1C

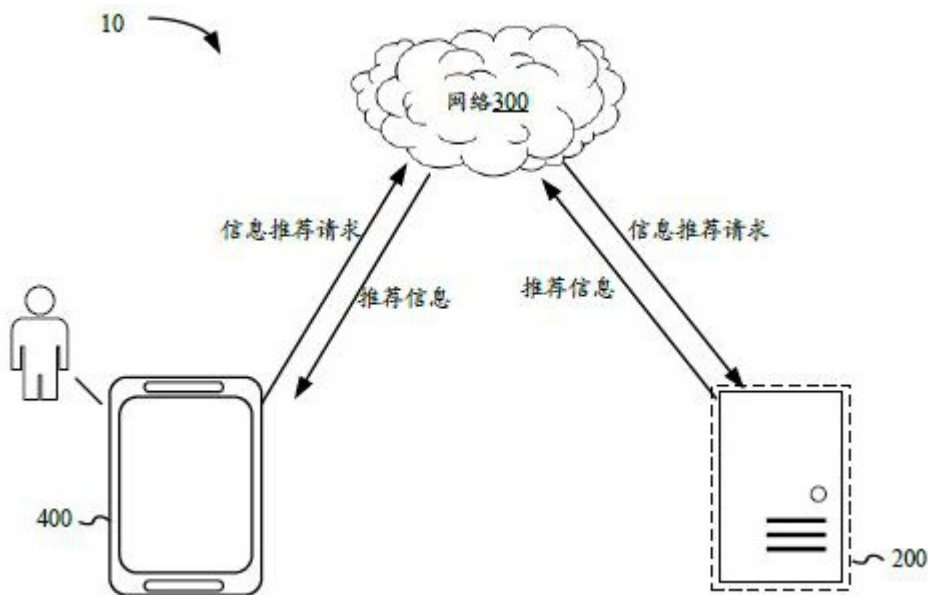


图2

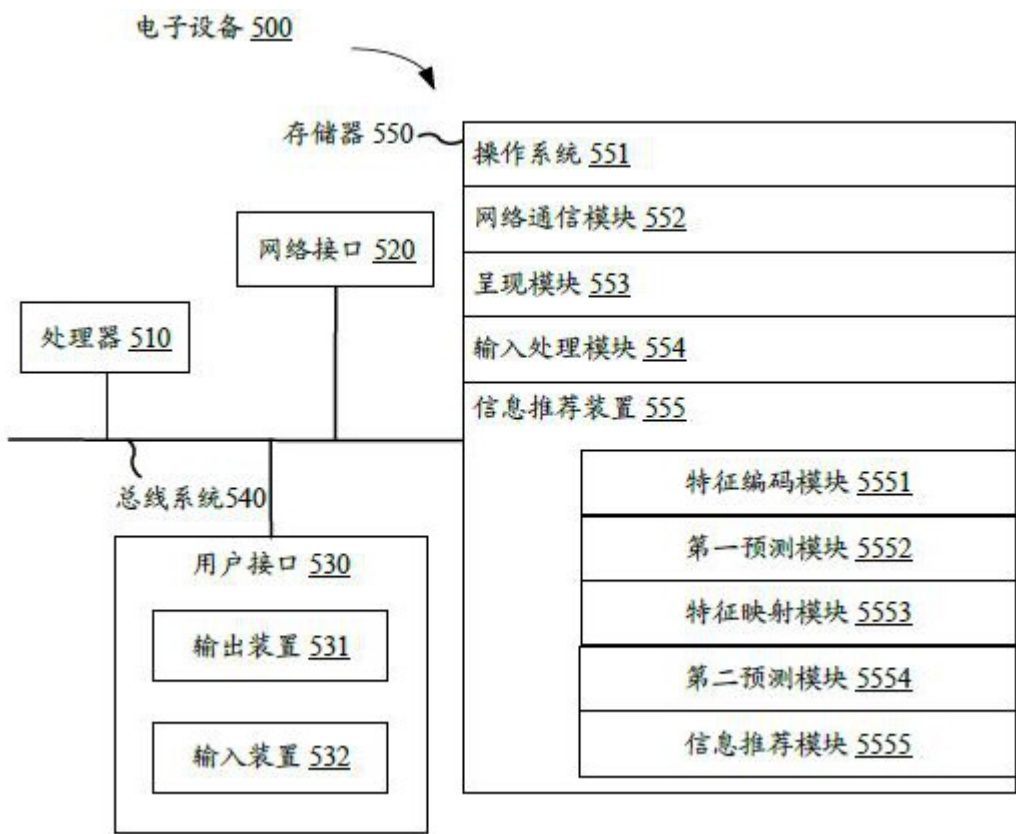


图3

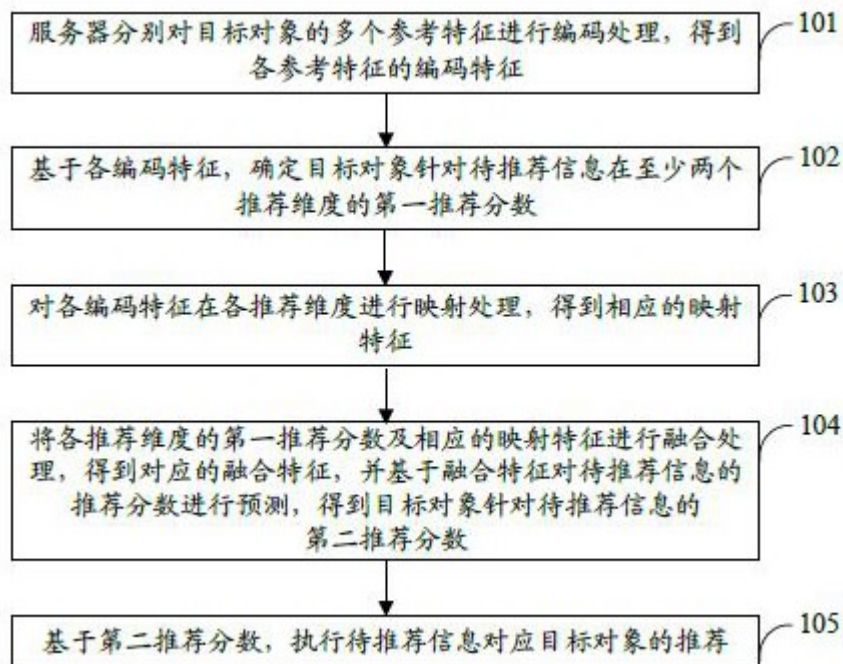


图4A

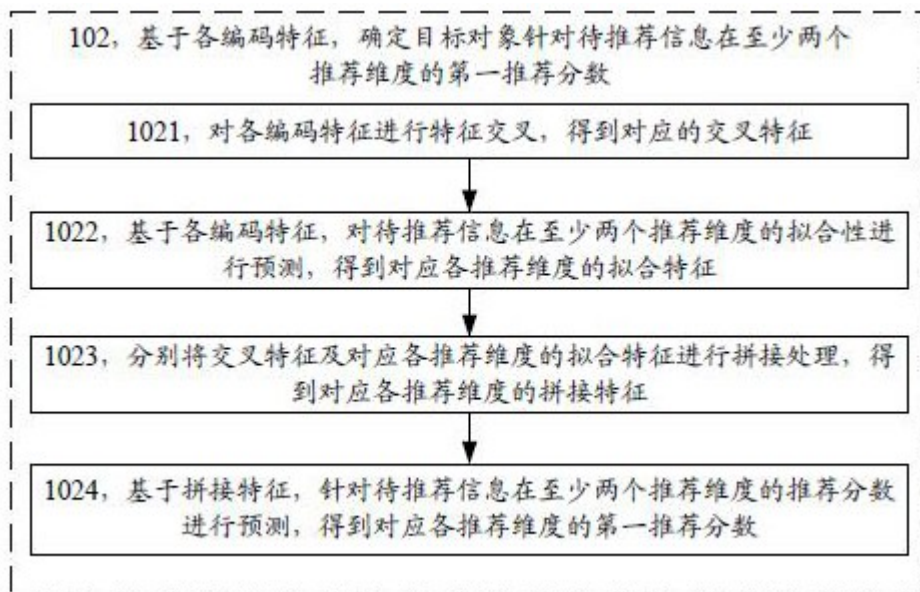


图4B

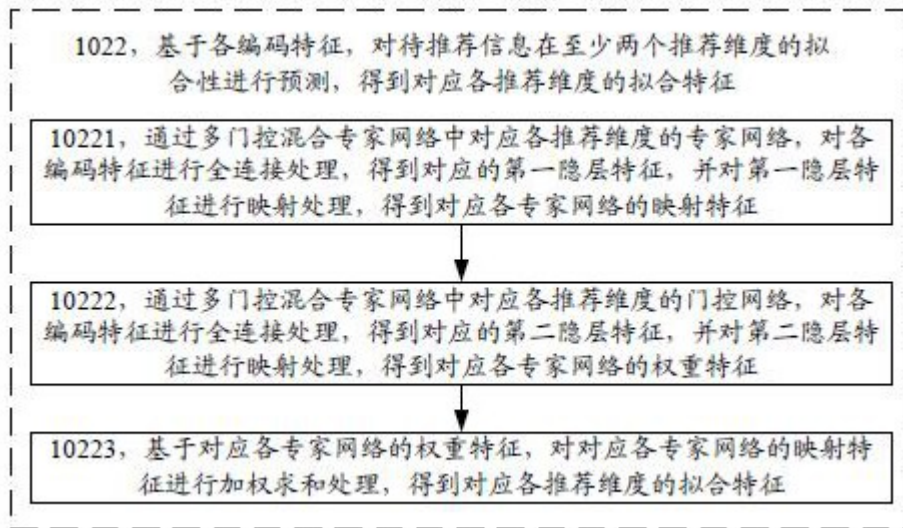


图4C

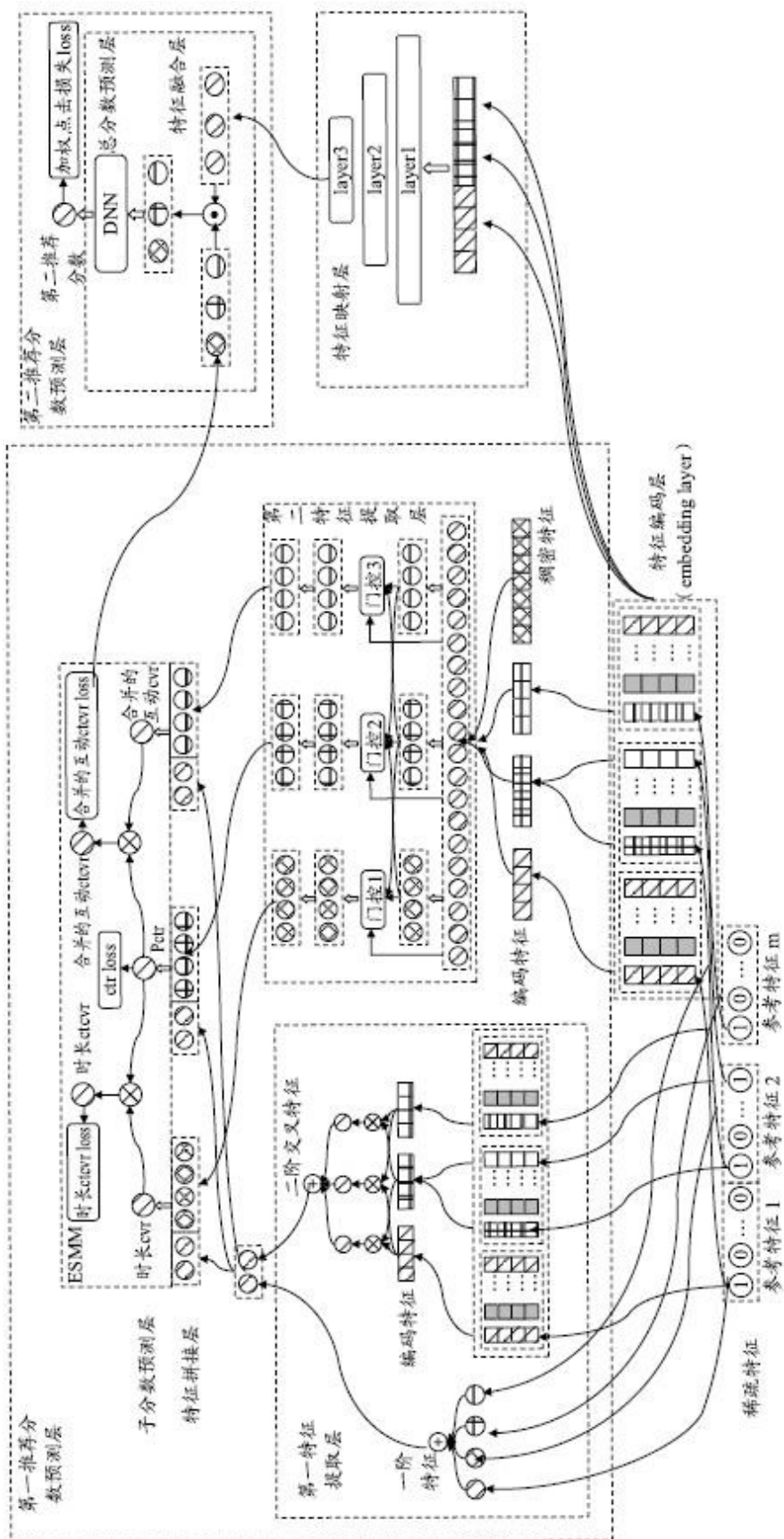


图5

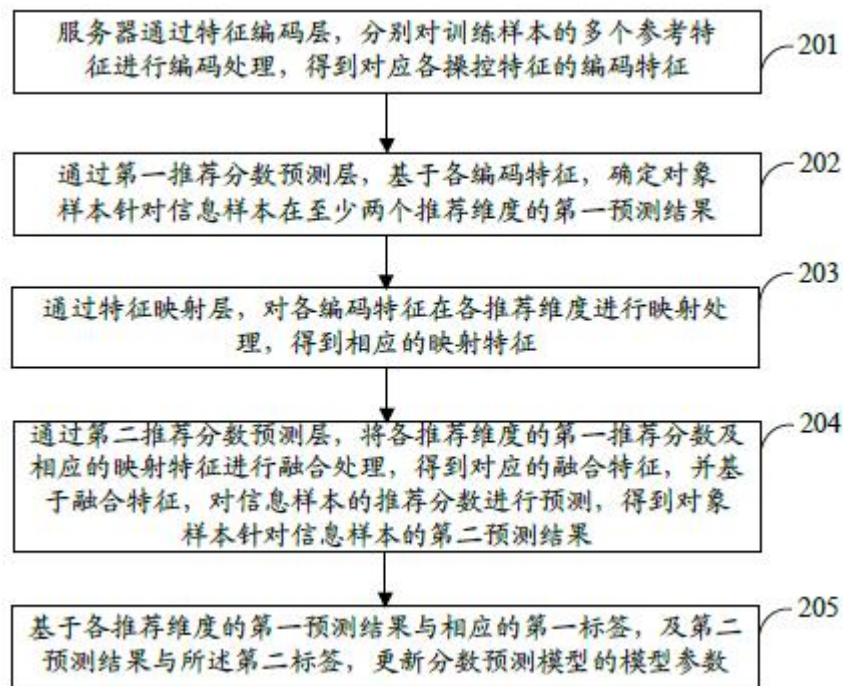


图6A

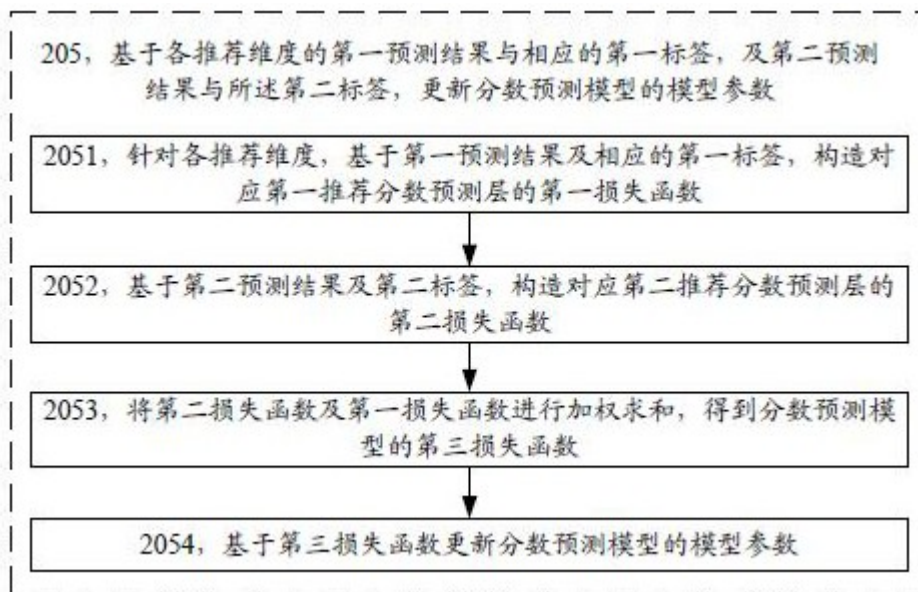


图6B

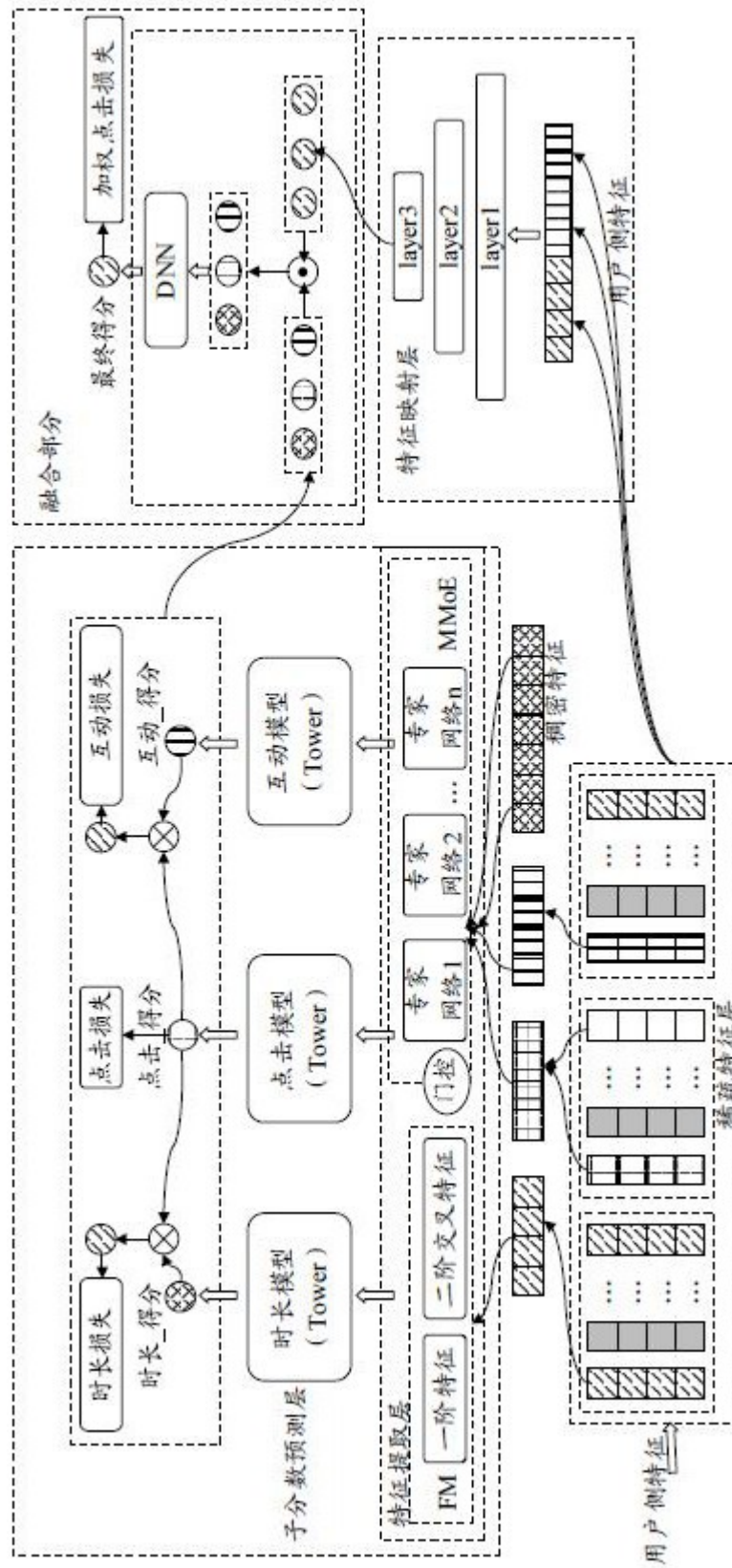


图7

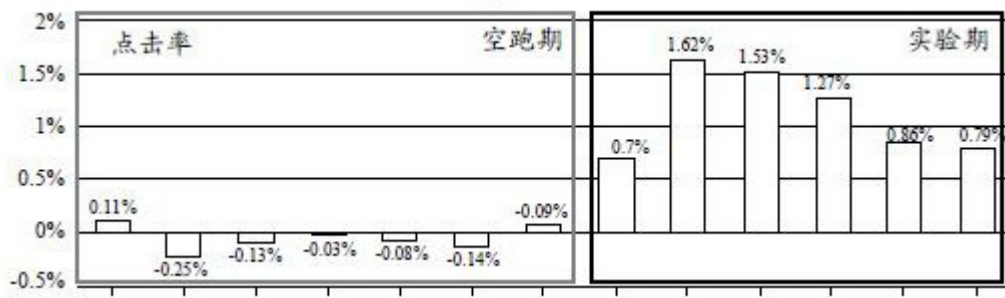


图8A

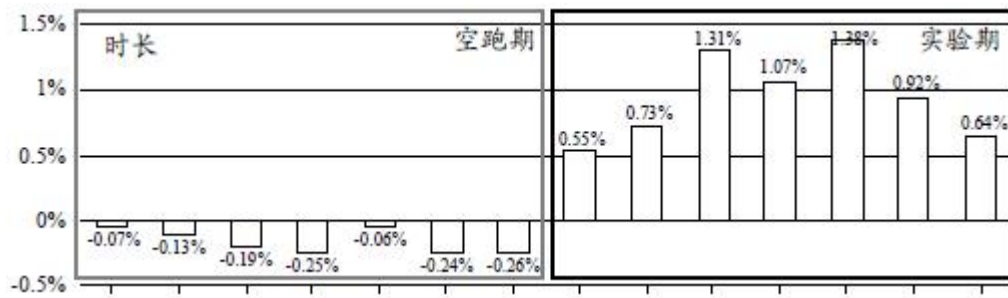


图8B

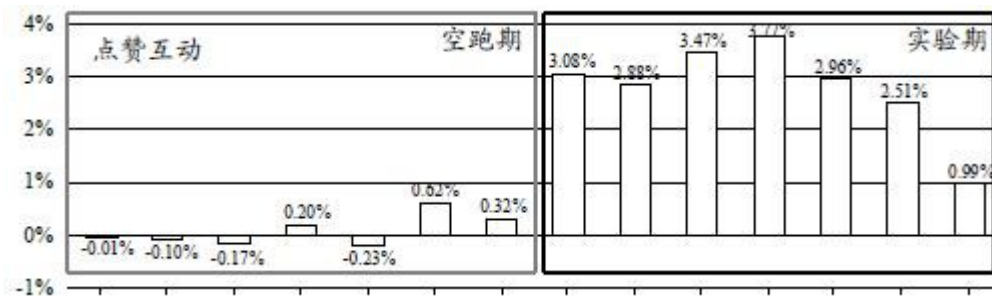


图8C