一种基于数据画像的双向就业推荐方法和装置

技术领域

本发明涉及推荐系统领域，具体涉及一种基于数据画像的双向就业推荐方法和装置。

背景技术

就业推荐系统的出现是为了帮助求职者和企业从海量繁杂的数据中快速而准确地获取所需的内容，主要算法有基于内容的算法和协同过滤算法。基于内容的算法是基于求职者的注册信息进行的推荐；而协同过滤算法是根据求职者的相似群体或者原有的岗位投递获取匹配的岗位信息，而后进行推荐。

本发明人发现，以上的所提的相关算法并没有考虑到以下两个问题：第一个问题，求职者对岗位的要求并不是一成不变的，它会随着时间的变化而变化；当变化发生时，基于原注册信息或原相似群体产生的推荐对于求职者而言就是一些冗余信息。第二个问题，人资的偏好会对求职者能否得到面试邀请产生一定的影响。人是有主观性的生命，这不可避免地将自己的偏好带入到工作中，导致以下的现象产生：即使岗位与求职者双方十分契合，但简历依旧可能过不了人资的筛选。

发明内容

本发明为了解决上述问题，提出了一种基于数据画像的双向就业推荐方法和装置，通过建立用户和岗位的数据画像，实现用户到岗位、以及岗位到用户双向的精准的个性化推荐，满足用户和企业的双向需求。

为了实现上述目的，本发明采用下述技术方案：

一种基于数据画像的双向就业推荐方法和装置，包括：

用户数据模块：用于处理和保存用户的基本信息，并将该信息导入到用户模块用于生成用户基本标签；

岗位数据模块：用于处理和保存岗位的基本信息，并将该信息导入到岗位模块用于生成岗位基本标签；

日志模块：用于保存和处理XML消息日志，同时统计和保存个人（包括用户和人资）的行为数据，并将该数据导入用户模块或岗位模块用于生成行为标签；

画像标签模块：用于定义画像标签的数据模型，定义在不同数据源情况下，相应标签的计算任务，并将相应的任务导入到用户模块或岗位模块；

用户模块：用于生成用户的基本标签和行为标签，进而获取用户画像，同时导入关联模块用于计算用户推荐列表；

岗位模块：用于生成岗位的基本标签和行为标签，进而获取岗位画像，同时导入关联模块用于计算岗位推荐列表；

关联模块：用于接收用户模块或岗位模块导入的数据画像，计算基本信息和行为信息的权重关系，然后用户和岗位画像的匹配度，并根据行为标签进行排序，生成推荐列表。

进一步，所述用户数据模块包括：

用户数据清理模块：用于对用户的结构信息（包括姓名、年龄、性别、工作年限、薪资、学历）进行清洗、去重，并利用注册用户的同属性均值对缺失值进行补充；同时对用户的非结构信息（包括自我评价、兴趣爱好、备注等短文本信息）进行分词和去停用词处理，获取关键词；

用户数据存储模块：用于保存用户的基本信息（结构信息和关键词），并将该数据导入用户模块用于基本标签的生成。

进一步，所述岗位数据模块包括：

岗位数据清理模块：用于对岗位的结构信息（包括对年龄、性别、工作年限、学历的要求以及薪资）进行清洗、去重，并利用同类型岗位的同属性均值对缺失值进行补充，获取岗位基本信息；并对岗位的非结构信息（包括岗位描述文本和所属企业的描述文本等）进行分词和去停用词处理，获取关键词；

岗位数据存储模块：用于保存岗位的基本信息（结构信息和关键词），并将该数据导入岗位模块用于基本标签的生成。

进一步，所述日志模块包括：

日志保存模块：用于暂时保存个人（包括用户和人资）通过浏览、点击、收藏、分享、简历投递（面试邀请）等操作产生的XML消息日志；

日志处理模块：用于处理和分析XML 消息日志，统计个人对各个标签的不同操作以及相应的次数、时长，保存为行为数据，并将该数据导入用户模块或岗位模块。

进一步，所述画像标签模块包括：

标签定义模块：用于定义标签的数据模型，以及标签的可解释性含义（如标签名或标签含义解释），所述标签包括年龄、性别、工作年限、薪资、学历、行业类型、企业类型、爱好、品质；

标签计算模块：根据不同的标签以及数据源，定义相应的计算任务，每一个计算任务中包含对数据的预处理，计算和整合操作。

进一步，所述用户模块包括：

用户基本标签模块：用于接收用户数据模块导入的数据，并根据画像标签模块导入的标签计算任务，生成用户的基本标签；

用户行为标签模块：用于接收日志模块导入的用户行为数据，并根据画像标签模块导入的标签计算任务，生成用户的行为标签；

用户画像模块：用于保存用户的基本标签和行为标签，即为用户画像。

进一步，所述岗位模块包括：

岗位基本标签模块：用于接收岗位数据模块导入的岗位数据，并根据画像标签模块导入的标签计算任务，生成岗位的基本标签；

人资行为标签模块：用于接收日志模块导入的人资行为信息，并根据画像标签模块导入的标签计算任务，生成人资行为标签；

岗位画像模块：用于保存岗位的基本标签和人资行为标签，即为岗位画像。

进一步，所述关联模块包括：

权值模块：用于对基本信息和行为信息分别进行归一化处理，而后计算权值，并将其分别与相应的基本信息和行为信息进行相乘，生成偏重不同的数据画像；

匹配模块：用于计算用户和岗位之间的匹配度，生成候选推荐列表；

推荐模块：根据行为标签对候选推荐列表进行排序，生成推荐列表。

一种基于数据画像的双向就业推荐方法和装置的业务流程构建过程，包括：

步骤201，处理和保存用户的信息；

步骤202，生成用户的基本标签和行为标签，进而获取用户画像；

步骤203，处理和保存岗位的信息；

步骤204，生成岗位的基本标签和行为标签，进而获取岗位画像；

步骤205，接收用户模块或岗位模块导入的数据画像，计算基本信息和行为信息的权值，而后计算用户和岗位画像的匹配度，并根据行为标签进行排序，生成推荐列表。

本发明的有益效果是：

1.本发明所属内容通过数据画像显示地展现用户和岗位的基本信息和偏好信息，强化了推荐结果的可解释性，降低了用户和岗位匹配度的计算量。

2.本发明通过分析人资的行为数据，挖掘岗位的偏好，实现对岗位的个性化推荐，降低求职者的投递风险，减少人资筛选简历的时间。

3.本发明考虑了基本信息和行为信息的占比对推荐结果的影响，通过计算基本信息和行为信息的权值，挖掘数据画像的兴趣偏向。

附图说明

图1是一种基于数据画像的双向就业推荐方法和装置的构成；

图2是一种基于数据画像的双向就业推荐方法和装置的业务流程构建过程。

具体实施方式通过

下面结合附图与实施例对本发明作进一步说明。

如图1所示，一种基于数据画像的双向就业推荐方法和装置，由用户数据模块101、岗位数据模块102，日志模块103，画像标签模块104，用户模块105，岗位模块106，关联模块107七部分组成。

用户数据模块101：主要负责处理和保存用户的基本信息，用户数据模块101通过用户数据清理模块1011、用户数据存储模块1012共同完成；其中，

用户数据清理模块1011：用于对用户的结构信息（包括姓名、年龄、性别、工作年限、薪资、学历）进行清洗、去重，并利用注册用户的同属性均值对缺失值进行补充；同时对用户的非结构信息（包括自我评价、兴趣爱好、备注等短文本信息）进行分词和去停用词处理，获取关键词；

用户数据存储模块1012：用于保存用户的基本信息（结构信息和关键词）。

岗位数据模块102：主要负责处理和保存岗位的基本信息，岗位数据模块102通过岗位数据清理模块1021、岗位数据存储模块1022共同完成；其中，

岗位数据清理模块1021：用于对岗位的结构信息（包括对年龄、性别、工作年限、学历的要求以及薪资）进行清洗、去重，并同类型岗位的同属性均值对缺失值进行补充，获取岗位基本信息；并对岗位的非结构信息（包括岗位描述文本和所属企业的描述文本等）进行分词和去停用词处理，获取关键词；

岗位数据存储模块1022：用于保存岗位的基本信息（结构信息和关键词）。

日志模块103：主要负责保存和处理XML消息日志，同时统计和保存个人（包括用户和人资）的行为数据，日志模块103通过日志保存模块1031、日志处理模块1032共同完成；其中，

日志保存模块1031：用于暂时保存个人（包括用户和人资）通过浏览、点击、收藏、分享、简历投递（面试邀请）等操作产生的XML消息日志；

日志处理模块1032：用于处理和分析XML 消息日志，统计个人对各个标签的不同操作以及相应的次数、时长，保存为行为数据，并将该数据导入用户模块或岗位模块。

画像标签模块104：主要负责定义画像标签的数据模型，定义在不同数据源情况下，相应标签的计算任务，画像标签模块104通过标签定义模块1041、标签计算模块1042共同完成；其中，

标签定义模块1041：用于定义标签的数据模型，以及标签的可解释性含义（如标签名或标签含义解释），所述标签包括年龄、性别、工作年限、薪资、学历、行业类型、企业类型、爱好、品质；

标签计算模块1042：根据不同的标签以及数据源，定义相应的计算任务，每一个计算任务中包含对数据的预处理，计算和整合操作。对于基本标签中的年龄、性别、工作年限、薪资、学历的属性，主要是直接获取用户或岗位的基本信息数据；对于基本标签中的行业类型、企业类型、爱好、品质的属性，主要是利用word2vec和支持向量机对用户或岗位的关键词进行相应的类别预测。对浏览、点击、收藏、分享、简历投递（面试邀请）等操作设置不同的权重；对于以产生的行为标签设置时间权重，该标签出现的时间越久远影响越弱。

用户模块105：主要负责生成用户的基本标签和行为标签，进而获取用户画像，用户模块105由用户基本标签模块1051、用户行为标签模块1052、用户画像模块1053共同完成；其中，

用户基本标签模块1051：用于接收用户数据模块101导入的数据，并根据画像标签模块104导入的标签计算任务，生成用户的基本标签；

用户行为标签模块1052：用于接收日志模块103导入的用户行为数据，并根据画像标签模块104导入的标签计算任务，生成用户的行为标签；

用户画像模块1053：用于保存用户的基本标签和行为标签，即为用户画像。

岗位模块106：主要负责生成岗位的基本标签和行为标签，进而获取岗位画像，岗位模块106通过岗位基本标签模块1061、岗位行为标签模块1062、岗位画像模块1063共同完成；其中，

岗位基本标签模块1061：用于接收岗位数据模块102导入的数据，并根据画像标签模块104导入的标签计算任务，生成岗位的基本标签；

人资行为标签模块1062：用于接收日志模块103导入的人资行为信息，并根据画像标签模块104导入的标签计算任务，生成人资行为标签；

岗位画像模块1063：用于保存岗位的基本标签和人资行为标签，即为岗位画像。

关联模块107：主要负责计算基本信息和行为信息的权值，而后计算用户和岗位画像的匹配度，并根据行为标签进行排序，生成推荐列表，关联模块107通过权值模块1071、匹配模块1072、推荐模块1073共同完成；其中，

权值模块1071：用于对基本信息和行为信息分别进行归一化处理，而后计算权值，并将其分别与相应的基本信息和行为信息进行相乘，生成偏重不同的数据画像；计算公式如下：

对基本信息进行归一化处理：

对行为信息进行归一化处理，处理后的行为信息表示为：

计算基本信息和行为信息的相似度:

若，表示基本信息和行为信息相同，则；

若，表示基本信息和行为信息存在包含不同的信息，则考虑基本信息和行为信息的信息量对数据画像的影响：

如果基本信息包含的信息量多于行为信息的，即，则，;

如果基本信息包含的信息量少于行为信息的，即，则，;

则基本信息为：，行为信息为

匹配模块1072：用于计算用户和岗位之间的匹配度，生成候选推荐列表；匹配度的计算公式如下：

利用欧式距离计算用户画像和岗位画像的相似度：

其中，表示岗位，表示用户；

则匹配度表示为。

推荐模块1073：根据用户或人资的权值较大的标签对候选推荐列表进行排序，生成推荐列表。‘

患者就诊路径智能规划方法和装置；

’

如图3所示，一种基于数据画像的双向就业推荐方法和装置的业务流程构建过程，包括以下步骤：

步骤201，处理和保存用户的信息；经1011用户数据清理模块对用户结构信息和非结构信息的处理，保存到1012用户数据存储模块，并发送到105用户模块用于生成基本标签；根据1031日志保存模块对用户日志的保存，经1032日志处理模块对其的处理和分析，并发送到105用户模块用于生成行为标签。

步骤202，接收101用户数据模块导入的数据，并根据104画像标签模块导入的标签计算任务，经1051用户基本标签模块生成用户的基本标签；接收103日志模块导入的用户行为数据，并104根据画像标签模块导入的标签计算任务，经1052用户行为标签模块生成用户的行为标签；经1053用户画像模块保存用户的基本标签和行为标签（即为用户画像），并发送到107关联模块用于生成推荐列表。

步骤203，处理和保存岗位的信息；经1021岗位数据清理模块对用户结构信息和非结构信息的处理，保存到1022岗位数据存储模块，并发送到106岗位模块用于生成基本标签；根据1031日志保存模块对人资日志的保存，经1032日志处理模块对其的处理和分析，并发送到106岗位模块用于生成行为标签。

步骤204，接收102岗位数据模块导入的用户数据，并根据104画像标签模块导入的标签计算任务，经1061岗位基本标签模块生成岗位的基本标签；接收103日志模块导入的人资行为数据，并104根据画像标签模块导入的标签计算任务，经1062人资行为标签模块生成人资的行为标签；经1063用户画像模块保存用户的基本标签和人资行为标签（即为岗位画像），并发送到107关联模块用于生成推荐列表。

步骤205，接收101用户数据模块导入的用户画像和102岗位数据模块导入的岗位画像，经1071权值模块、1072匹配模块、1073推荐模块生成推荐列表。

上述虽然结合附图对本发明的具体实施方式进行了描述，但并非对本发明保护范围的限制，所属领域技术人员应该明白，在本发明的技术方案的基础上，本领域技术人员不需要付出创造性劳动即可做出的各种修改或变形仍在本发明的保护范围以内。

图1

患者就医监控

诊疗资源监控

历史诊疗资源占用评估

---计算---

图2

1.一种基于数据画像的双向就业推荐的方法和装置，其特征是，包括：

用户数据模块：用于处理和保存用户的基本信息，并将该信息导入到用户模块用于生成用户基本标签；

岗位数据模块：用于处理和保存岗位的基本信息，并将该信息导入到岗位模块用于生成岗位基本标签；

日志模块：用于保存和处理XML消息日志，同时统计和保存个人（包括用户和人资）的行为数据，并将该数据导入用户模块或岗位模块用于生成行为标签；

画像标签模块：用于定义画像标签的数据模型，定义在不同数据源情况下，相应标签的计算任务，并将相应的任务导入到用户模块或岗位模块；

用户模块：用于生成用户的基本标签和行为标签，进而获取用户画像，同时导入关联模块用于计算用户推荐列表；

岗位模块：用于生成岗位的基本标签和行为标签，进而获取岗位画像，同时导入关联模块用于计算岗位推荐列表；

关联模块：用于接收用户模块或岗位模块导入的数据画像，计算基本信息和行为信息的权重关系，然后用户和岗位画像的匹配度，并根据行为标签进行排序，生成推荐列表。

2.如权利要求1所述的一种基于数据画像的双向就业推荐的方法和装置，其特征是，所述用户数据模块包括：

用户数据清理模块：用于对用户的结构信息（包括姓名、年龄、性别、工作年限、薪资、学历）进行清洗、去重，并利用注册用户的同属性均值对缺失值进行补充；同时对用户的非结构信息（包括自我评价、兴趣爱好、备注等短文本信息）进行分词和去停用词处理，获取关键词；

用户数据存储模块：用于保存用户的基本信息（结构信息和关键词），并将该数据导入用户模块用于基本标签的生成。

3.如权利要求1所述的一种基于数据画像的双向就业推荐的方法和装置，其特征是，所述岗位数据模块包括：

岗位数据清理模块：用于对岗位的结构信息（包括对年龄、性别、工作年限、学历的要求以及薪资）进行清洗、去重，并利用同类型岗位的同属性均值对缺失值进行补充，获取岗位基本信息；并对岗位的非结构信息（包括岗位描述文本和所属企业的描述文本等）进行分词和去停用词处理，获取关键词；

岗位数据存储模块：用于保存岗位的基本信息（结构信息和关键词），并将该数据导入岗位模块用于基本标签的生成。

4. 如权利要求1所述的一种基于数据画像的双向就业推荐的方法和装置，其特征是，所述日志模块包括：

日志保存模块：用于暂时保存个人（包括用户和人资）通过浏览、点击、收藏、分享等操作产生的XML消息日志；

日志处理模块：用于处理和分析XML 消息日志，统计个人对各个标签的不同操作以及相应的次数、时长，保存为行为数据，并将该数据导入用户模块或岗位模块。

5.如权利要求1所述的一种基于数据画像的双向就业推荐的方法和装置，其特征是，所述画像标签模块包括：

标签定义模块：用于定义标签的数据模型，以及标签的可解释性含义（如标签名或标签含义解释），所述标签包括年龄、性别、工作年限、薪资、学历、行业类型、企业类型、爱好、品质；

标签计算模块：根据不同的标签以及数据源，定义相应的计算任务，每一个计算任务中包含对数据的预处理，计算和整合操作。

6.如权利要求1所述的一种基于数据画像的双向就业推荐的方法和装置，其特征是，所述用户模块包括：

用户基本标签模块：用于接收用户数据模块导入的数据，并根据画像标签模块导入的标签计算任务，生成用户的基本标签；

用户行为标签模块：用于接收日志模块导入的用户行为数据，并根据画像标签模块导入的标签计算任务，生成用户的行为标签；

用户画像模块：用于保存用户的基本标签和行为标签，即为用户画像。

7.如权利要求1所述的一种基于数据画像的双向就业推荐的方法和装置，其特征是，所述岗位模块包括：

岗位基本标签模块：用于接收岗位数据模块导入的岗位数据，并根据画像标签模块导入的标签计算任务，生成岗位的基本标签；

人资行为标签模块：用于接收日志模块导入的人资行为信息，并根据画像标签模块导入的标签计算任务，生成人资行为标签；

岗位画像模块：用于保存岗位的基本标签和人资行为标签，即为岗位画像。

8.如权利要求1所述的一种基于数据画像的双向就业推荐的方法和装置，其特征是，所述关联模块包括：

权值模块：用于对基本信息和行为信息分别进行归一化处理，而后计算权值，并将其分别与相应的基本信息和行为信息进行相乘，生成偏重不同的数据画像

匹配模块：用于计算用户和岗位之间的匹配度，生成候选推荐列表；

推荐模块：根据行为标签对候选推荐列表进行排序，生成推荐列表。

9.如权利要求1所述的一种智能的就业推荐方法和装置，其特征是，包括：

步骤201，处理和保存用户的信息；

步骤202，生成用户的基本标签和行为标签，进而获取用户画像；

步骤203，处理和保存岗位的信息；

步骤204，生成岗位的基本标签和行为标签，进而获取岗位画像；

步骤205，接收用户模块或岗位模块导入的数据画像，计算基本信息和行为信息的权值，而后计算用户和岗位画像的匹配度，并根据行为标签进行排序，生成推荐列表。

本发明公开了一种基于数据画像的双向就业推荐的方法和装置，其特征在于，首先处理和保存用户和岗位的基本信息；同时分析和处理用户和人资的日志信息，获取行为信息；然后通过定义画像标签计算任务，分别生成用户和岗位的数据画像；最后计算基本信息和行为信息的权值，而后计算用户和岗位的匹配度，并利用行为标签进行排序，生成推荐列表。本发明基于数据画像显示地展现用户和岗位的基本信息和偏好信息，强化了推荐结果的可解释性；考虑人资的行为数据，挖掘岗位的偏好，实现对岗位的个性化推荐，降低求职者的投递风险；同时考虑了基本信息和行为信息的占比对推荐结果的影响，通过计算基本信息和行为信息的权值，挖掘数据画像的兴趣偏向

