

基于关系图谱的人岗关系研究

王晓萍¹, 郭梦洁^{2,3}, 岳靖雯^{2,3}

1. 上海市经济和信息化委员会信息中心, 上海 200125; 2. 复旦大学计算机科学技术学院, 上海 201203;
3. 上海市数据科学重点实验室, 上海 201203

摘要

利用现有的数据进一步挖掘分析并帮助干部组织工作, 是一个既有挑战又具有潜力的方向。针对干部信息数据的特点, 使用基于关系图谱的人岗相宜研判方法分析领导班子, 通过整合干部履历表以及人员基本信息库中的多源信息, 构建关系图谱; 将基于网络表示学习算法提取的关系图谱中的节点及关系等信息作为特征并输入分类模型, 实现人岗关系研判。实验结果表明, 基于关系图谱的方法可以很好地捕获人员和岗位之间的复杂关系信息, 准确地判断人岗关系。

关键词

人岗关系分析; 关系图谱; 网络表示学习; 大数据

中图分类号: TP311

文献标识码: A

doi: 10.11959/j.issn.2096-0271.2020059

Research on person-position relationship based on relation graph

WANG Xiaoping¹, GUO Mengjie^{2,3}, YUE Jingwen^{2,3}

1. Shanghai Municipal Commission of Economy and Informatization, Information Center, Shanghai 200125, China
2. School of Computer Science, Fudan University, Shanghai 201203, China
3. Shanghai Key Laboratory of Data Science, Shanghai 201203, China

Abstract

Utilizing the existing data to further analyze and help leaders organize their work is a potential and challenging direction. According to the characteristics of leader information data, the leadership team was analyzed using the person-position relationship judgment method based on relation graph, a relation graph was built by integrating the leader resume and the multi-source information from the database. Then the information such as nodes and relationships in the relation graph extracted by the network representation learning method was used as features to input into the classification model. By using the proposed model, the relationship between people and positions can be inferred. The experimental results show that the method based on relation graph can well capture the complex relationship between people and positions, and can accurately judge the person-position relationship.

Key words

analysis on person-position relationship, relation graph, network representation learning, big data

1 引言

干部综合研判是一项系统性很强的工作,需要收集掌握各类信息资料,对人员进行全方位、立体化的研判。判断是否人岗相宜,找出领导班子的最优搭配方案是提高干部综合水平的一个有效方法。如何借助信息化手段开展干部综合研判,帮助组织部门更好地管理、任用和选拔干部,是一个具有挑战的问题。大数据为决策提供了越来越多的支持,通过对组织系统中的数据资源的整合,可以有效打破数据壁垒,推进数据资源的综合开发利用,提高组织工作的科学化水平。现实中的许多复杂关系可以用图(或网络)的形式进行描述和分析处理,图(或网络)中的节点表示数据实体,边表示关系,指示节点对之间的关联关系。图结构的数据已经在大量机器学习任务中发挥了重要作用,如节点分类^[1]、链接预测^[2]、商品推荐^[3]等。

在干部信息管理过程中,最初解决的是干部个人信息电子化的问题,初步实现了全部数据可查询、可统计、可输出。但在实际应用中,单纯的数据存储、数据查询已不能满足日益发展的业务需求,需要引入新的数据分析方法,推动技术与业务的深度融合,进一步促进干部组织工作的顺利开展。在领导班子搭建过程中,需要考虑众多因素,比如哪些人有过共事关系,哪些人是同学或者校友,哪些人曾经先后在同一部门工作,哪些人有过相关部门的工作经验,哪些人一起参加过培训班,哪些人已经在同一岗位任职多年(或部门名称变化但是岗位职责没变)。在实际工作过程中,直接通过干部履历表人工分析或搜索这些相应的工作经历及学习培训经历,是一件非常耗时费力的工作。

为应对上述困难,本文基于网络表示学习方法对图中的节点进行特征分析和提取,从而构建人员关系图谱,实现人员与岗位的匹配推荐,支持人岗相宜分析。基于关系图谱和网络表示学习方法实现部门成员的人岗相宜分析,分析干部履历中工作过的部门,以及这些工作部门与领导班子其他成员之间的交集关系、干部履历中的属性,从而判断人员和岗位是否匹配,为组织部门搭配领导班子提供辅助思路。

本文的主要工作包括以下几个方面。

(1)从异构数据源中提取数据,进行数据清洗、数据整合,构建人员关系网络。此外,将人员的特征属性等形成多维度标签。

(2)人岗关系的发现:利用网络表示学习方法进行人员特征学习表示,提取节点(人员、岗位等)的特征和网络结构、关系等。本文采用异质网络(heterogeneous network)表示学习算法metapath2vec^[4]对构建的多源异质的人员关系网络进行特征表示。

(3)基于提取的网络特征进行人员和岗位关系预测,使用XGBoost^[5]分类模型分析人岗关系的匹配程度。

(4)将由(2)(3)组成的两阶段方法称为基于网络表示学习的人岗关系预测(network representation learning for person-position relationship identification, NRL4PPR)方法,实验结果表明了NRL4PPR方法的有效性。

(5)为了体现NRL4PPR方法的可伸缩性,对一个大规模异质网络数据集进行网络关系预测。实验结果进一步表明,NRL4PPR方法可以处理多种来源、多种类型的大规模数据。

(6)在上述工作的基础上,给出一个人岗相宜研判系统架构,为判断在职干部是否人岗相宜提供支持。

2 相关工作

2.1 人事关系研究

早期人事关系方面的研究主要通过问卷调查方式开展^[6-7],其规模非常有限,且缺少灵活高效的分析方法。Chien C F^[8]等人使用关联分析方法发现企业职位与候选人的关系,用于企业招聘候选人;Jantan H等人^[9]利用分类模型评估员工的工作表现;Li L Y等人^[10]采用循环神经网络对员工离职后的工作岗位进行预测。这些方法虽然在人岗关系研究方面取得了一定的成效,但是没有充分利用员工之间的关系。

近年来,图数据分析技术因为利用了节点之间的关系以及图的结构特征,引起了研究者广泛的研究兴趣,且被引入人事关系研究。Liu Y等人^[11]运用多源多任务学习模型融合了分布在多个社交网络上的信息,以预测用户的职业轨迹。胥皇^[12]引入属性图作为行为表示,提出基于个人工作记录数据的职业轨迹属性图提取算法,并针对职业流动行为问题,构建了人才流动图。杜鹏程等人^[13]利用社会网络图谱,可视化分析了国内人力资源的情况,掌握了人力资源领域的研究热点。上述研究体现了图数据挖掘技术在人力资源领域的有效应用,即根据数据信息构建关系网络,从网络中挖掘更多的价值,但是上述方法没有考虑到异质网络中不同边表示的关系类型的差异(例如有的边表示毕业于同一院校,有的边表示参加过同一培训班),没有提取网络中的语义特征。

2.2 网络表示学习

网络表示学习的目标是将网络中的节

点映射到低维向量空间,并尽可能保留网络的拓扑结构及节点特征等信息。网络表示学习将网络数据同后续机器学习任务链接起来,可用于系统推荐、节点分类、链路预测等。

随着深度学习技术的出现,研究者提出了基于深度神经网络的表示学习模型。例如DeepWalk^[14]和node2vec^[15]通过在网络上进行随机游走,捕获每个节点附近的邻居,从而学习节点的特征向量。LINE分别捕获1阶和2阶邻域关系,以学习网络的局部和全局信息^[16]。GCN通过在网络上定义传播规则来为每个节点融合其邻居特征信息,最终得到每个节点的向量表示^[1]。同质网络上的表示学习方法难以区分不同类型的关系,无法捕获丰富的语义信息,因此研究者提出异质网络的表示学习方法。具有代表性的是采用元路径(meta-path)指导随机游走^[4],以更好地获取异质网络中更加复杂丰富的结构和语义信息。例如metapath2vec^[4]算法基于元路径的随机游走,实现根据不同类型节点生成含有网络语义的异质邻居。此外,扩展Skip-Gram模型^[17]促进了对位置和语义接近的节点的建模。人员关系图谱中包含人员、岗位等多种类型的节点和边的关系,引入元路径来捕捉语义信息有助于提升人岗关系分析的有效性。

3 人岗关系分析方法

本节利用网络表示学习技术进行人岗关系分析,主要包括构建成员关系网络,以及基于网络表示学习抽取节点和关系特征,从而分析人岗关系。

3.1 构建成员关系网络

干部信息系统经过了多次人员信息核

定,是准确度较高的数据来源。关于干部个人信息及单位信息,涉及的表格包括基本情况表、学历表、工作经历表、年度考核表、家庭成员及社会关系表等。此外,干部的个人信息、单位信息还包括与人员行为相关的数据,如人员参与的会议、因公务出差的情况、参与的项目等行为数据。

首先对上述原始数据进行数据预处理和融合,将分散在多个数据源的数据进行映射,如存在曾用名的人员与其工作经历的映射等。然后将人员、岗位、工作单位、毕业院校、项目等设置为图的节点,并且将人员的出生年月等看作人员类型节点的属性;同时,构建人员与岗位、人员与人员、人员与工作单位、人员与毕业院校、人员与会议及项目等行为之间、岗位与岗位之间的关系(即边)。加入边的属性描述作为辅助信息,例如张三和复旦大学的关系可能有本科毕业(学历)、研究生毕业(学历)等不同的属性,加入属性描述能更有效地对该节点进行特征表示。最后形成一个人员关系图谱,因为图中的节点和关系类型有多种,所以该图谱是一个异质信息网络^[18],并且具有节点属性和边属性。

人员关系采用一种基于图的数据结构来表示,它由节点(实体)和边(关系)组成,每个节点代表一个实体,边表示实体与实体之间的关系,实现了相关数据的打通,构建了包含多数据源的人员关系图谱(如图1所示)。考虑到数据的时效性,需定期对数据进行更新,通过增删节点(如工作单位变动)与更新节点属性(如新增奖惩)等对关系网络进行维护。

3.2 基于网络表示学习的人岗关系分析

传统的特征工程方法对人工的依赖程度高,网络表示学习方法已被广泛用于将

网络中的节点转化为特征向量表示。如前文所述,网络表示学习通过分析节点之间的拓扑结构关系等,为每个节点生成一个特征向量。该特征向量蕴涵了节点自身的特征信息与节点在网络中的关联信息,可直接应用于多种下游任务,如链接预测与节点分类。

定义1 异质网络^[18]:异质网络是一种包含了多种类型的节点和边的网络,用 $G=(V,E)$ 表示。其中, V 表示图中节点的集合, $E\subseteq V\times V$ 表示所有节点两两之间的关系,不同类型的节点对之间的关系构成了多种类型的边。 A 和 R 分别表示节点类型和边类型的集合,若 $|A|+|R|>2$,则称之为异质网络,否则,称之为同质网络。

本文构建的人员关系异质网络的一个简单示例如图2所示。该示例中显示了3种类型的节点,以及节点间已存在的关系(图2中实线箭头)。进行人岗关系分析时要预测人员和岗位之间的关系(虚线箭头)。

定义2 元路径^[19]:元路径表示异质网络中节点间的一种复合关系,记为 $P=V_1\rightarrow V_2\rightarrow\cdots\rightarrow V_M$, P 描述了节点类型 V_1 到 V_M 存在的一种复合关系。

不同的元路径携带不同的语义信息,例如对于图2中的异质网络示例,可以定义元路径“人员 \rightarrow 院校 \rightarrow 人员”表示存在不同人员毕业于同一院校,元路径“岗位 \rightarrow 人员 \rightarrow 院校 \rightarrow 人员 \rightarrow 岗位”表示两个岗位存在毕业于同一院校的人员(如图3所示)。

定义3 异质网络表示学习(也称为异质网络嵌入)^[19]:异质网络表示学习的目的是使异质网络中的节点学习低维稠密的向量表示。即给定一个异质网络 $G=(V,E)$,异质网络表示学习是对网络上的每一个节点 $v\in V$ 学习一个实数向量 $X_v\in R^d$,其中向量的维度 d 远小于节点总数。

异质网络表示学习的目标是使学习的低维表示能最大限度地保留网络中多种类

型的节点和边的关系。异质网络表示学习算法可以处理不同规模的数据（如包含千万个节点的大型复杂网络），其考虑了节点的不同类型以及关系的不同类型，可以很好地处理不同来源、不同类别的异质数据，从而捕获数据中丰富的语义关系信息。

本文使用基于元路径的异质网络表示学习方法metapath2vec^[4]进行人岗关系分析，具体步骤如下。

(1) 节点序列生成

通过随机游走生成由图中节点构成的若干序列。首先选择一个节点作为起始节点，在图上进行随机游走，即以一定概率从当前节点的邻居节点中抽取一个节点进行跳转，并加入序列，再从该邻居节点重复抽取-跳转的过程，直到序列达到预定的长度。在这一步骤中，应选择不同的节点作为起始节点，从而生成多条节点序列。

(2) 节点序列训练

节点序列蕴涵了图中的关联结构信息，从这些序列中生成节点的特征表示。这里应用了分布假说，即一个节点的含义可由它的上下文节点（即节点序列中前后若干节点）反映出来。基于Skip-Gram方法，给定一个节点作为中心节点，利用其特征表示来预测上下文节点。节点的特征表示向量是随机初始化的，在训练过程中，特征向量不断优化调整，直至可较为准确地预测上下文节点。此时，节点特征向量已经蕴涵了图中的关联结构信息。

(3) 引入元路径

考虑到该关系网络的边和节点都是多种类型的，基于元路径的异质网络节点特征表示学习方法metapath2vec是一种基于元路径的随机游走方法，可针对各种类型的节点生成含有网络语义的异质邻居。给定元路径 $\mathcal{P}=V_1 \rightarrow V_2 \rightarrow \dots \rightarrow V_{t+1}$ ，从类型为 V_k 的节点 v_k 出发，随机游走，下一步只会访

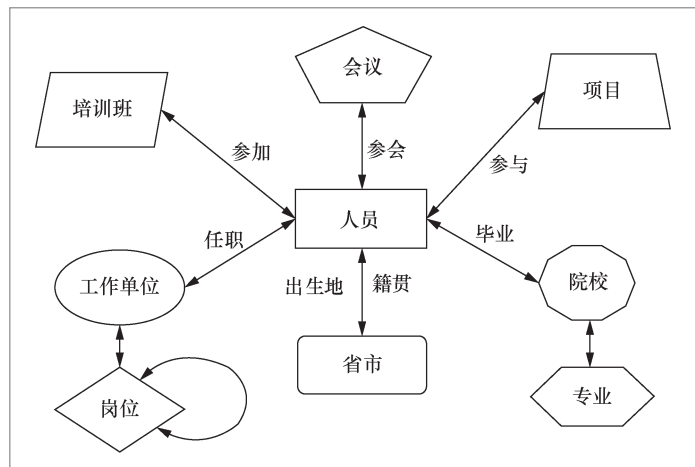


图1 人员关系图谱模式

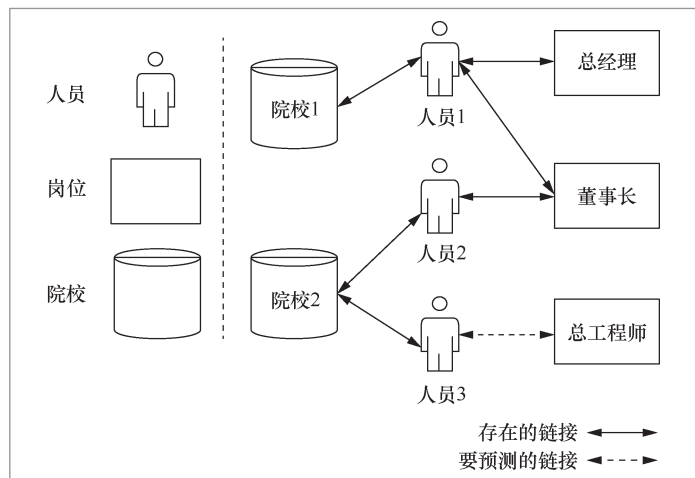


图2 人员关系异质网络的简单示例

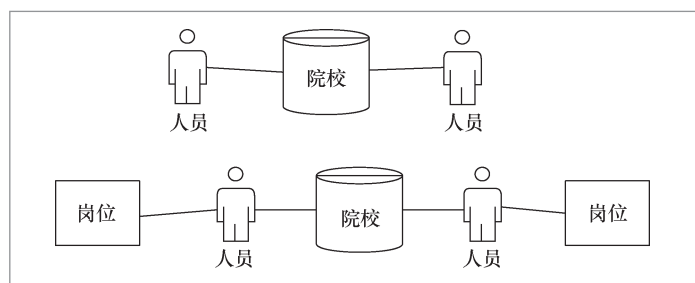


图3 异质网络示例中定义的两条元路径

问类型为 V_{k+1} 的节点。元路径指导的随机游走在网络上生成节点序列，将节点序列输入模型中进行训练。

(4) 人岗关系分析

由于人员与岗位均为图中的节点, 利用上述过程得到人员与岗位的特征表示向量, 便可进行人岗关系分析。人岗关系分析的核心为计算人员节点与岗位节点的特征向量的相似度。例如人岗相宜程度可通过计算相应向量的余弦相似度或内积进行量化。对于岗位推荐, 可以将其视为链接预测任务, 通过现有的人岗关系数据, 在特征向量的基础上训练分类模型, 该模型可判断网络中哪些人岗存在潜在的链接, 即人岗适配。本文实验采用XGBoost分类器, 将人员岗位等特征向量作为特征, 以预测人员和岗位、人员和机构之间是否存在关系。

4 数据与实验

本节首先描述数据集及其处理, 然后进行实验, 最后对实验结果进行分析。

4.1 数据描述

为了说明本文提出的基于网络的节点特征表示学习方法在人岗适配中的有效性, 笔者在真实数据上进行了实验。考虑到数据的可公开性, 选取来自中国经济金融研究数据库(China Stock Market & Accounting Research Database, CAMAR)的上市公司等数据构建了一个具有人员S(staff)、岗位O(occupation)、上市公司E(enterprise)、行业D(domain)、非上市机构I(institution)和毕业院校U(university)等节点的异质网络。该异质网络中包含这些节点间的多种关系, 如人员与公司的雇佣关系, 人员与人员之间的亲属关系, 人员从哪所大学毕业, 公司所属的领域等。具体说明如下。

数据包括2018年开始在沪深股市上市的公司的任职董事、监事以及高管等人员, 并以这些人员为中心, 收集他们的亲属、在非上市公司的同事、毕业院校等信息。最终提取出6种类型的节点: 人员S(如蒋某、杨某)、岗位O(如总经理、总工程师)、上市公司E(如深圳发展银行股份有限公司、万科企业股份有限公司)、行业D(如医药制造业、零售业)、非上市公司或机构I(如复旦大学经济学院、中华全国工商业联合会)和毕业院校U(如复旦大学、北京大学), 以及6种关系: 上市公司任职关系 $S \xleftarrow{\text{employed}} E$ (人员在上市公司任职)、人员所属岗位 $S \xleftarrow{\text{position}} O$ (人员担任某岗位)、亲属关系 $S \xleftarrow{\text{relative}} S$ (人员和另一位人员之间有亲属关系)、毕业院校关系 $S \xleftarrow{\text{graduate}} U$ (人员毕业于该院校)、非上市公司任职关系 $S \xleftarrow{\text{work}} I$ (人员在非上市公司或机构任职)、所属行业关系 $S \xleftarrow{\text{business}} D$ (上市公司的所属行业)。在基于该数据集构建的异质网络中, 节点与关系的数据统计见表1。另外, 本数据集考虑了人员同名的问题, 每一个人员都具有独立ID, 以便区分。

4.2 方法对比与参数设置

人员岗位关系预测实验采用3条元路径($O \rightarrow S \rightarrow E \rightarrow S \rightarrow O$, 表示两个岗位存在在一家上市公司任职的人员; $O \rightarrow S \rightarrow O$, 表示两个岗位存在共同的人员; $O \rightarrow S \rightarrow U \rightarrow S \rightarrow O$, 表示两个岗位存在毕业于同一院校的人员)进行实验。

人员机构关系预测实验选用3条元路径($I \rightarrow S \rightarrow E \rightarrow S \rightarrow I$, 表示两个机构之间存在在一家上市公司任职的人员; $I \rightarrow S \rightarrow I$, 表示两个机构之间存在共同的人员; $I \rightarrow S \rightarrow U \rightarrow S \rightarrow I$, 表示两个机构之间存在毕业于同一院校的人员)进行实验。

实验中采用DeepWalk和metapath2vec两种算法, DeepWalk是一种同质网络表示学习算法, 其在图上进行随机游走, 基于Skip-Gram模型得到节点向量表示。对于DeepWalk和metapath2vec, 将每个节点的游走次数设为100, 游走的最大节点序列长度设为100, 最终的向量表示维度设为64。

4.3 实验结果

4.3.1 人员岗位关系预测

对于数据原图, 去除30%岗位与人员的连边, 将剩下的节点和边作为训练集; 将所有被去除的边作为正样本, 在每个正样本中随机选取5条不存在的边作为负样本, 构建测试集。例如, 对于真实边(人员1, 岗位1), 可以采样虚假边(人员1, 岗位2)作为负样本。评价指标采用准确率(accuracy)和AUC值。采用DeepWalk和基于3条不同元路径(OSES0、OSO、OSUS0)的metapath2vec进行网络表示学习。在得到节点的表征后, 使用XGBoost分类器对以上数据集进行实验, 将训练集的比例分别设置为20%、40%、60%和80%。对人员和机构之间的关系进行推断分析, 实验结果见表2。将使用DeepWalk+XGBoost的预测方法记为DWPPR。

从表2可以看出, 相比于不考虑人员岗位的异质信息的DeepWalk, metapath2vec在不同训练比例中均能取得更好的实验结果。这展示了捕获异质语义信息的重要性, 突出了NRL4PPR在人岗关系分析中的优越性。

此外, 选择OSO元路径的metapath2vec基本取得了最好的实验结果(除了在20%的训练比例上比OSES0元路径的准确率低0.001, 但实际应用中训练比例都超过20%), 主要原因是该元路径很好地考虑了

表1 节点与关系的数据统计

节点类型	数量/个	关系类型	数量/个
S	22 536	S $\xleftarrow{\text{employed}}$ E	7 590
O	30	S $\xleftarrow{\text{position}}$ O	4 144
E	1 646	S $\xleftarrow{\text{relative}}$ S	882
D	73	S $\xleftarrow{\text{graduate}}$ U	1 326
I	8 427	S $\xleftarrow{\text{work}}$ I	55 669
U	377	E $\xleftarrow{\text{business}}$ D	1 728

表2 人员岗位关系预测的实验结果

评价指标	训练集	DWPPR	NRL4PPR		
			OSES0	OSO	OSUS0
准确率	20%	0.841	0.869	0.868	0.868
	40%	0.838	0.871	0.872	0.871
	60%	0.837	0.871	0.872	0.871
	80%	0.831	0.874	0.876	0.875
AUC	20%	0.793	0.847	0.849	0.844
	40%	0.794	0.836	0.849	0.843
	60%	0.783	0.842	0.847	0.844
	80%	0.766	0.835	0.838	0.829

人员与岗位的对应关系, 而且长的元路径在捕获网络链接结构时用处不大^[20]。不同元路径对应的实验结果差别不大, 说明元路径可以捕获网络的异质关系信息, 且选用元路径的异质网络表示学习方法在人岗关系分析中优于不考虑异质语义的同质网络表示学习算法DeepWalk。

4.3.2 人员机构关系预测

对于数据原图, 去除30%机构与人员的连边, 将剩下的节点和边作为训练集; 将所有被去除的边作为正样本, 每个正样本同样随机选取5条不存在的边作为负样本, 构建测试集。评价指标也采用准确率和AUC值。采用DeepWalk和基于3条不同元路径(ISESI、ISI、ISUSI)的metapath2vec进行网络表示学习。之后

同样使用XGBoost分类器对以上数据集进行实验,同样分别将训练集的比例设置为20%、40%、60%和80%。对人员和机构之间的关系进行推断分析,实验结果见表3。

从表3可以看出,metapath2vec在不同元路径的实验结果均优于DeepWalk。这主要是因为metapath2vec通过元路径捕获了异质网络中丰富的语义信息,而DeepWalk未考虑网络的异质性,将人员和机构视为同种类型的节点。

此外,可以看到选择ISI作为元路径的metapath2vec获得了最好的实验结果,因为该元路径很好地考虑了人员与机构的交互。而采用ISESI元路径并未取得较好的实验结果,这是因为该元路径中引入了上市公司,而数据中人员与上市公司的链接较为稀疏,导致产生的节点序列长度较短,影响了模型的效果。同样,人员与大学的链接也较为稀疏,因此采用ISUSI的元路径也未取得较好的实验结果。但是从总体上来说,选用元路径的异质网络表示学习方法在人员机构关系分析中优于不考虑语义的DeepWalk。

4.4 扩展实验

为了展示本文采用的方法的伸缩性和扩展性,进一步使用AMinerCS^[21]数据

集进行实验。该数据集包括1 693 531名作者A、3 194 405篇论文P和3 883种会议或期刊C,以及它们之间的关系,包括论文在哪个会议或期刊上发表($P \xrightarrow{\text{publish}} C$)、论文的作者($P \xrightarrow{\text{written}} A$)。节点与关系的数据统计见表4。

采用基于元路径C→P→C的metapath2vec对AMinerCS数据集进行网络表示学习。训练集和测试集的划分和之前的实验设置相同。之后同样使用XGBoost分类器对该数据集进行实验,对论文和会议之间的关系进行预测,实验结果见表5。

从表5可以看出,本文采用的NRL4PPR方法在大规模数据集上也取得了较高的准确率和AUC,体现了其良好的伸缩性和扩展性。本文采用的方法不仅可以应用在人员公司数据集上进行人岗关系匹配,也可以应用于千万级别数目的大规模数据集上进行有效的关系预测。

5 人岗相宜研判系统

人岗相宜研判系统建立在上述工作的基础上,为判断在职干部是否人岗相宜提供了支持。该系统主要通过数据融合、特征分析和人岗推荐,实现对干部的数据集成和分析推荐。其主要架构如图4所示,包括数据层、模型层和应用层3个层次。

(1) 数据层

收集到系统所需的人员及其岗位等信息之后,进行数据脱敏。在人员关系图存储方面,使用Neo4j作为数据存储数据库。

(2) 模型层

通过网络表示学习对数据层得到的关系图谱进行表征学习,提取网络中丰富的特征和关系信息,网络中每个节点都能得到一个向量表示。根据特征向量可对人岗关系进行分析,构成人员深度特征标签,并

表3 人员机构关系预测的实验结果

评价指标	训练集	DWPPR	NRL4PPR		
			ISESI	ISI	ISUSI
准确率	20%	0.864	0.872	0.892	0.870
	40%	0.869	0.876	0.900	0.871
	60%	0.869	0.879	0.901	0.873
	80%	0.867	0.877	0.902	0.873
AUC	20%	0.761	0.814	0.871	0.802
	40%	0.777	0.830	0.884	0.809
	60%	0.787	0.833	0.890	0.818
	80%	0.788	0.832	0.894	0.822

将其作为传统人工标注标签的补充。

(3) 应用层

应用层主要由基础查询、人岗推荐模块组成。

- 基础查询模块实现了人员的基本信息查询,即根据数据库中存储的数据进行信息查询。

- 人岗推荐模块实现了干部与岗位之间的匹配,即根据构建的关系图谱,基于网络表示学习的人岗关系发现方法进行人岗推荐。

本文使用上海市经济和信息化委员会在编部门及人员数据对人岗相宜研判系统进行验证。目前该委员会共有35个处级(含副处级)处室部门以及17个事业单位(处级),涉及公务员294名和事业单位在编人员1 274名。将上述相关信息进行匿名脱敏处理后,构建异质关系图谱,并输入研判系统,主要涉及的关系如图1所示。当需要寻求某个岗位的合适候选人时,可以输入岗位信息,然后人岗相宜研判系统可以基于人岗关系预测算法推荐与该岗位相匹配的人员信息,根据元路径的语义定义,可以获得曾经在同一部门或同一岗位工作过、毕业于同一个专业、参加过同一个培训、参加过同一个项目等的匹配结果,关系匹配相似度越高的人员,就越有可能成为候选人。如以某处室的副处级岗位进行候选人推荐,通过对比现任副处长的关系图谱,顺利找到3~4名候选人,经相关部门专家对候选人与岗位匹配结果进行评估,发现匹配度较高,验证了人岗相宜研判系统的有效性。

6 结束语

本文利用构建的关系图谱和基于网络表示学习的人岗推荐方法,分析人员履历

表4 AMinerCS 异质网络数据统计

节点类型	数量/个	关系类型	数量/个
A	1 693 531	$P \xrightarrow{\text{publish}} C$	3 194 405
P	3 194 405	$P \xrightarrow{\text{written}} A$	9 323 739
C	3 883		

表5 论文会议关系预测的实验结果

评价指标	准确率				AUC			
	20%	40%	60%	80%	20%	40%	60%	80%
训练集	0.853	0.855	0.855	0.854	0.800	0.801	0.802	0.803
NRL4PPR (CPC)	0.853	0.855	0.855	0.854	0.800	0.801	0.802	0.803

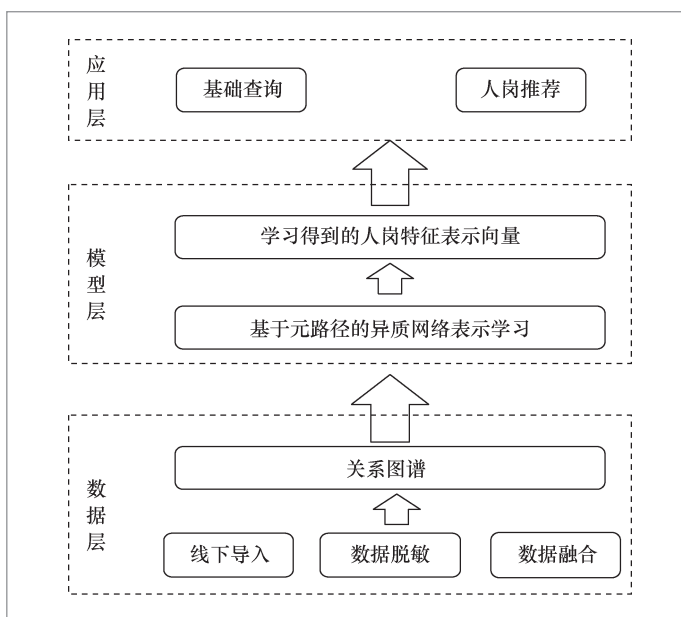


图4 系统整体架构

中工作过的单位/机构,以及这些工作单位/机构与其他成员之间的交集关系、履历中的属性。实验结果表明,本文方法可以很好地判断人岗是否相宜,为组织部门搭班提供了辅助思路。

参考文献:

- [1] KIPF T N, WELLING M. Semi-supervised classification with graph convolutional

- networks[J]. arXiv preprint, 2016, arXiv: 1609.02907.
- [2] ZHANG M, CHEN Y. Link prediction based on graph neural networks[C]// The 32nd Conference on Neural Information Processing Systems. [S.l.:s.n.], 2018: 5165–5175.
- [3] FAN S H, ZHU J X, HAN X T, et al. Metapath-guided heterogeneous graph neural network for intent recommendation[C]// The 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM Press, 2019: 2478–2486.
- [4] DONG Y X, CHAWLA N V, SWAMI A. metapath2vec: scalable representation learning for heterogeneous networks[C]// The 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2017: 135–144.
- [5] CHEN T Q, GUESTRIN C. XGBoost: a scalable tree boosting system[C]// The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2016: 785–794.
- [6] NG T W H, SORENSEN K L, EBY L T, et al. Determinants of job mobility: a theoretical integration and extension[J]. *Journal of Occupational and Organizational Psychology*, 2007, 80(3): 363–386.
- [7] SCHEIN E H. *Career dynamics: matching individual and organizational needs*[M]. Boston: Addison Wesley Publishing Company, 1978.
- [8] CHIEN C F, CHEN L F. Data mining to improve personnel selection and enhance human capital: a case study in high-technology industry[J]. *Expert Systems with applications*, 2008, 34(1): 280–290.
- [9] JANTAN H, HAMDAN A R, OTHMAN Z A. Knowledge discovery techniques for talent forecasting in human resource application[J]. *World Academy of Science, Engineering and Technology*, 2009, 50: 775–783.
- [10] LI L Y, JING H, TONG H H, et al. Nemo: next career move prediction with contextual embedding[C]// The 26th International Conference on World Wide Web Companion. [S.l.:s.n.], 2017: 505–513.
- [11] LIU Y, ZHANG L M, NIE L Q, et al. Fortune teller: predicting your career path[C]// The 30th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2016: 201–207.
- [12] 胥皇. 基于属性图挖掘的职业流动行为研究[D]. 西安: 西北工业大学, 2018.
- XU H. Understanding career mobility behavior based on attributed graph mining[D]. Xi'an: Northwestern Polytechnical University, 2018.
- [13] 杜鹏程, 吴婷, 王成城. 科技人力资源研究领域的知识图谱分析[J]. *中国科技论坛*, 2013, 1(8): 83–89.
- DU P C, WU T, WANG C C. The knowledge mapping analysis of the research areas of human resources for science and technology[J]. *Forum on Science and Technology in China*, 2013, 1(8): 83–89.
- [14] PEROZZI B, AL-RFOU R, SKIENA S. DeepWalk: online learning of social representations[C]// The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2014: 701–710.
- [15] GROVER A, LESKOVEC J. node2vec: scalable feature learning for networks[C]// The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2016: 855–864.
- [16] TANG J, QU M, WANG M, et al. LINE: large-scale information network embedding[C]// The 24th International Conference on World Wide Web. [S.l.:s.n.], 2015: 1067–1077.
- [17] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality[C]// The 26th Conference on Neural Information Processing Systems. [S.l.:s.n.], 2013: 3111–3119.

- [18] SUN Y Z, YU Y T, HAN J W. Ranking-based clustering of heterogeneous information networks with star network schema[C]// The 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2009: 797-806.
- [19] LU Y F, SHI C, HU L M, et al. Relation structure-aware heterogeneous information network embedding[C]// The 33th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2019: 4456-4463.
- [20] SUN Y Z, HAN J W, YAN X F, et al. PathSim: meta path-based top- k similarity search in heterogeneous information networks[J]. Proceedings of the VLDB Endowment, 2011, 4(11): 992-1003.
- [21] TANG J, ZHANG J, YAO L M, et al. ArnetMiner: extraction and mining of academic social networks[C]// The 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2008: 990-998.

作者简介



王晓萍 (1977-), 上海市经济和信息化委员会信息中心高级工程师, 主要研究方向为大数据、政府信息化。



郭梦洁 (1994-), 复旦大学计算机科学技术学院、上海市数据科学重点实验室硕士生, 主要研究方向为大数据、网络表示学习方法。



岳婧雯 (1997-), 复旦大学计算机科学技术学院、上海市数据科学重点实验室硕士生, 主要研究方向为大数据、网络表示学习方法。

收稿日期: 2020-04-02

基金项目: 国家自然科学基金资助项目 (No.1636207, No.61671157); 上海市经济和信息化委员会市级部门预算资金项目 (No.201859971904)

Foundation Items: The National Natural Science Foundation of China (No.1636207, No.61671157), The Foundation of Shanghai Municipal Commission of Economy and Informatization (No.201859971904)