

# 网络表示学习

陈维政, 张岩, 李晓明

北京大学信息科学技术学院 北京 100871

## 摘要

以Facebook、Twitter、微信和微博为代表的大型在线社会网络不断发展,产生了海量体现网络结构的数据。采用机器学习技术对网络数据进行分析的一个重要问题是如何对数据进行表示。首先介绍了网络表示学习的研究背景和相关定义。然后按照算法类别,介绍了当前5类主要的网络表示学习算法,特别地,对基于深度学习的网络表示学习技术进行了详细的介绍。之后讨论了网络表示学习的评测方法和应用场景。最后,探讨了网络表示学习的研究前景。

## 关键词

网络;表示学习;深度学习

doi: 10.11959/j.issn.2096-0271.2015025

## *Network Representation Learning*

**Chen Weizheng, Zhang Yan, Li Xiaoming**

School of Electronic Engineering and Computer Science, Peking University, Beijing 100871, China

### *Abstract*

Along with the constant growth of massive online social networks such as Facebook, Twitter, Weixin and Weibo, a tremendous amount of network data sets are generated. How to represent the data is an important aspect when we apply machine learning techniques to analyze network data sets. Firstly, the research background was introduced and the definitions of NRL (network representation learning) were related. According to the categories of different algorithms, five kinds of primary NRL algorithms were introduced. Particularly, a detailed introduction to NRL algorithms based deep learning techniques was given emphatically. Then the evaluation methods and application scenarios of NRL were discussed. Finally, the research prospect of NRL in the future was discussed.

### *Key words*

network, representation learning, deep learning

## 1 引言

在互联网兴起之前，社会网络分析一直是社会科学、心理学和人类学等专业的研究领域，数据规模一直都相对较小。伴随着互联网的快速普及，以Facebook、Twitter、微信、微博为代表的大型在线社会网络不断发展，这些网络动辄拥有数千万乃至数亿用户，产生的数据极其丰富，结构极其复杂，成为人类社会生活的一种真实写照。同时，学术论文引用网络、移动通信网络、维基百科乃至整个互联网等具有网络结构的数据也得到了学术界广泛的研究。网络数据挖掘已经成为当前计算机科学的重要研究领域，每年在KDD、WSDM和ICDM等国际会议上都会发表大量相关的研究论文。

机器学习算法为相关研究提供了重要的模型和工具。如何对网络数据进行合理表示以作为机器学习算法的输入，是机器学习算法应用在网络数据分析中首先要面对的问题，从2013年开始举办的ICLR (International Conference on Learning Representations) 表明，表示学习已经成为学术界关心的重要问题。

网络数据最大的特点在于样本点之间存在着链接关系，这表明网络中样本点之间并非完全独立。除了节点间的链接关系，网络中节点自身也可能含有信息，比如互联网中网页节点对应的文本信息，这些特性使得网络表示学习需要考虑更多的因

素。传统的基于谱嵌入、最优化、概率生成模型等框架的网络表示算法已经不能适应大数据时代对网络表示学习问题在算法效率和精度方面的更高要求。近年来，基于深度神经网络的算法在特征学习上获得了极大的进展，给语音识别、图像识别、自然语言处理等领域带来了一种新气象。就自然语言处理而言，近两年以word2vec模型为代表的基于深度学习的词向量表示模型，掀起了一股表示学习的研究热潮。这些深度学习模型也启发了网络表示学习的研究，近两年的相关工作已经展现出这一方向的广阔前景。

## 2 相关定义

俗话说“巧妇难为无米之炊”，再强大的机器学习算法也需要数据进行支持。在同样的数据集上，针对同一个任务，由于特征的不同，同一个算法的结果也可能会有天壤之别。图1展示了应用机器学习算法的过程，首先从数据中提取有价值的信息，然后把数据表示成特征向量，进而采用机器学习算法完成相关任务。由于特征的选择和设计对算法结果的决定性作用，很多数据挖掘方面的研究工作沦为了特征工程，即把工作重心放到了针对特定的数据由人工设计出有价值的特征。因为数据的规模越来越大，复杂性越来越高，特征工程需要耗费大量的人力，而且需要特征设计者具备专业的领域知识，这与应用机器学习算法的初衷是相违背的。这种状况表

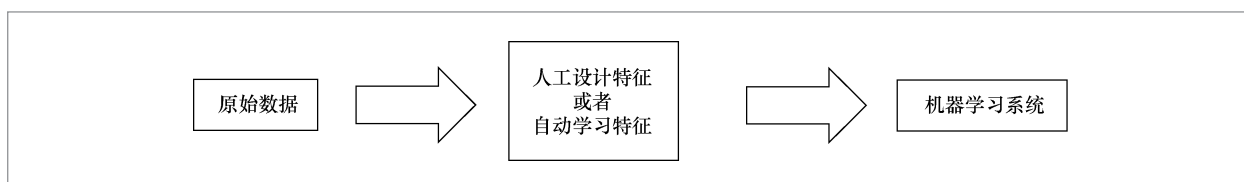


图1 应用机器学习算法的流程

明传统的机器学习算法严重依赖于特征的设计,而不能从原始数据中分辨出有价值的信息。

表示学习又称作特征学习,是机器学习领域中的一个重要研究问题,它的目标是自动学习一个从原始输入数据到新的特征表示的变换,使得新的特征表示可以有效应用在各种机器学习任务中,从而把人从繁琐的特征工程中解放出来。根据训练数据的区别,可以把表示学习算法分为如下两类。

- 监督表示学习 (supervised representation learning): 是指从已标注数据中学习数据的新的特征表示。比如处理分类或者回归任务的多层神经网络,其中的隐藏层可以作为输入数据的新的特征。参考文献[1]提出的有监督字典学习也是一类典型的有监督表示学习算法。

- 无监督表示学习 (unsupervised representation learning): 是指从未标注数据中学习数据的新的特征表示。无监督表示学习算法通常也是一种降维算法,用来从高维输入数据中发现有意义的低维特征表示。经典的代表性算法包括主成分分析、局部线性嵌入<sup>[2]</sup>、独立成分分析<sup>[3]</sup>和无监督字典学习<sup>[4]</sup>等。

参考文献[5]对表示学习的评价准则和主要方法进行了详细的综述,但对网络表示学习基本没有涉及,而且上述的表示学习算法通常都无法直接应用到网络数据上。在一个网络中,相连的节点之间通常存在着依赖关系。以网页分类任务为例,每个网页都可以认为是一个文档,只考虑文档的文本内容信息虽然也可以对网页进行分类,但却忽视了网页之间的链接关系。因为在现实世界中,网络中通常具有同质性,相似的节点更有可能存在着联系,直观上说就是“人以类聚,物以群分”以及“近朱者赤,近墨者黑”,这表明有可能通过网络中的链接信息获得更好的节点(这里指一

个网页)特征表示。参考文献[6]提出了通过网络表示学习把节点表示为向量的3种好处:

- 可以直接利用得到的节点向量表示作为机器学习算法的输入,避免针对网络数据重新设计新的机器学习算法;

- 网络中节点的距离、乘积等需要定量计算的概念不容易给出明确的定义,通过把节点表示成向量,可以在向量空间中直接进行各种运算;

- 在大规模网络数据中,节点之间的链接关系可能会非常复杂而不易观察,但是通过在低维向量空间中进行可视化分析展示,可以很直观地观察节点之间的关系。

为了给出网络表示学习的定义,下面首先介绍一些基本的概念。

- 节点 (vertex, node): 节点是网络中的一个功能个体,引文网络中的一篇文章、社交网络中的一个用户,都可以看作一个节点。

- 边 (edge): 边是用来刻画两个节点之间关系的,可能具有方向性,称为有向边,比如引文网络中的引用关系和邮件网络中的发送接收关系;也可能不具备方向性,称为无向边,比如Facebook中的好友关系。

- 网络 (network): 网络是对关系数据的刻画,定义网络 $G=(V,E)$ , $V$ 是 $G$ 的节点集合, $E$ 是 $G$ 的边集合。

- 邻接矩阵 (adjacency matrix): 网络 $G=(V,E)$ 对应的邻接矩阵 $A$ 是 $|V| \times |V|$ 的,邻接矩阵直接给出了网络的矩阵表示,如果 $(v_i, v_j) \in E$ ,那么 $A_{ij}=1$ ,否则 $A_{ij}=0$ 。

- 相似度矩阵 (similarity matrix): 网络 $G=(V,E)$ 对应的相似度矩阵 $W$ 是 $|V| \times |V|$ 的,如果没有额外的信息,可以通过把邻接矩阵 $A$ 中每一行进行归一化得到 $W$ 矩阵, $W$ 矩阵可以作为计算PageRank值时用到的转移概率矩阵。如果有额外的信

息, 比如在引文网络中, 可以用两篇论文的文档相似度作为节点之间的相似度。相似度矩阵有时也被称为亲和度矩阵 (affinity matrix)。

- 度数矩阵 (degree matrix): 网络  $G=(V, E)$  对应的度数矩阵  $D$  是  $|V| \times |V|$  的对角矩阵, 如果  $i=j$ , 那么  $D_{ij}=i$  的度数, 否则  $D_{ij}=0$ 。

- 拉普拉斯矩阵 (Laplacian matrix): 网络  $G=(V, E)$  对应的拉普拉斯矩阵  $L$  满足  $L = D - A$ 。

- 特征矩阵: 网络  $G=(V, E)$  只给出网络的链接关系, 而网络中的每个节点  $v$  可能拥有其他属性, 比如社会网络中的用户除了好友链接关系外, 自身还拥有标签数据, 在论文引用网络中, 每个文章节点还对应着自身的文本内容。定义  $G$  的特征矩阵为  $X$ ,  $X$  是  $|V| \times m$  维的,  $m$  是节点属性的特征空间大小。  $X$  通常是一个高度稀疏的矩阵, 传统的表示学习算法围绕在如何对  $X$  有效降维以获得数据点的低维表示方面, 而没有考虑网络的显式链接关系。

- 信息网络: 如果网络  $G=(V, E)$  对应的特征矩阵  $X$  是非空的, 那么  $G$  是一个信息网络。

借鉴参考文献[7]中对网络表示学习的定义并进行扩展, 给出如下定义。

**定义1** (网络表示学习) 给定网络  $G=(V, E)$ ,  $G$  对应的节点特征矩阵是  $X$ , 对任意节点  $v \in V$ , 学习低维向量表示  $r_v \in R^k$ ,  $r_v$  是一个稠密的实数向量, 并且满足  $k$  远小于  $|V|$ 。

定义1并不限定网络  $G$  的方向性, 即无论  $G$  中的边是有向的还是无向的, 都会予以考虑。定义1也不对网络  $G$  中的节点类型做出限制, 如果  $G$  中只含有一种类型的节点, 那么  $G$  是一个同构网络; 如果  $G$  中的节点属于不同的类型, 那么  $G$  是一个异构网络。由于针对异构网络表示学习的工作相对较少, 在下文中, 如果没有特别指出, 提到的

网络通常是指同构网络。特征矩阵  $X$  也可以是一个空矩阵, 由于隐私性的要求, 有时无法获得网络除了结构之外的数据, 此时只有网络中显式的链接信息, 节点本身不能提供更多的信息, 那么网络  $G$  是一个纯网络。反之, 如果  $G$  是非空矩阵, 那么网络  $G$  有结构之外的信息,  $G$  是一个信息网络。互联网就是一个典型的信息网络, 考虑到在线社会网络中存在大量的UGC (user generated content, 用户原创内容), 在线社会网络也可以构建为信息网络。

下文将针对上述各种情况下的网络表示学习进行介绍。直观上, 有两种依据可以对不同的算法进行分类: 一种是参考表示学习的分类, 即将所有的网络表示学习算法分为有监督的网络表示学习和无监督的网络表示学习; 另一种是根据输入数据的不同进行划分, 比如按照网络的方向性、是否是异构网络等性质。然而这两种划分依据并不合适, 因为当前的网络表示学习算法的主要区别在于算法类型, 同一算法类型下的算法框架都是相似的, 因此第3节将按照算法类型的区别对相关研究工作进行分类整理。

## 3 网络表示学习方法介绍

### 3.1 基于谱方法的网络表示学习

从广义上看, 谱方法是指利用输入数据矩阵的谱 (比如特征值和特征向量, 奇异值和奇异向量) 的一类算法的统称<sup>[7]</sup>。针对网络表示学习, 这个矩阵由特定的算法设计决定, 可能是相似度矩阵、拉普拉斯矩阵等类型。谱方法常用来获得数据的低维表示<sup>[8]</sup>, 比如经典的PCA (principal components analysis, 主成分分析) 算法中就是对样本的协方差矩阵选取特征向

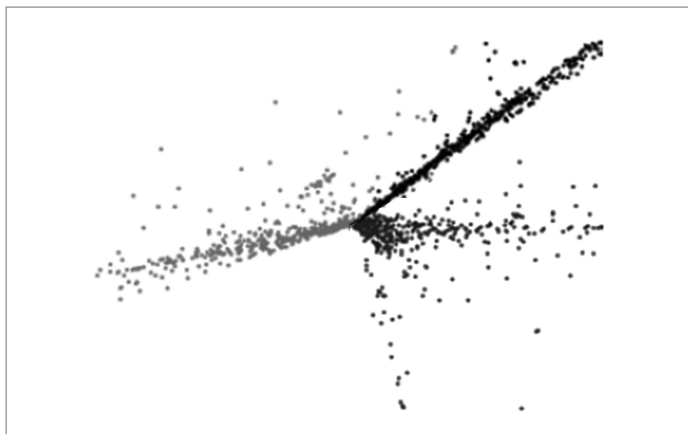


图2 DGE 算法的降维效果展示<sup>[6]</sup>

量进行降维。虽然可以把网络表示成邻接矩阵,作为PCA或者SVD(singular value decomposition, 奇异值分解)的输入以获得节点的低维表示,但是由于缺乏节点内在的信息,通常这种表示的质量较差<sup>[9]</sup>。基于谱方法的网络表示学习只考虑了结构信息,难以直接对信息网络进行应用,这里对几个代表性算法进行介绍。

参考文献[2]提出LLE(locally linear embedding)是一种非线性降维算法。LLE算法认为每一个数据点都可以由其近邻点的线性加权组合构造得到。LLE算法的输入是一个邻接矩阵<sup>[10]</sup>,然后计算出每个节点的局部重建权值矩阵,最后推导出一个特征值分解问题,进而计算节点的低维表示。

参考文献[11]提出的LaplacianEigenmaps算法的直观思想是希望相邻的节点在降维后的空间中尽可能地接近。LaplacianEigenmaps算法可以反映出数据内在的流形结构。LaplacianEigenmaps算法的输入也是邻接矩阵,与LLE算法不同的是LaplacianEigenmaps算法最终选取拉普拉斯矩阵的最小 $k$ 个非零特征值对应的特征向量作为学习到的网络表示。

LLE算法和LaplacianEigenmaps算法都只能处理无向网络,但现实中很多

网络(如网页链接网络)都是有方向的,Chen等基于随机游走的思想提出了DGE(directed graph embedding)算法<sup>[6]</sup>,DGE算法可以处理有向或无向网络,如果无向网络是连通的,DGE算法此时等价于LaplacianEigenmaps算法。图2给出了DGE算法把WebKB数据映射到二维空间的可视化效果,WebKB数据集包括康奈尔、德克萨斯和威斯康辛3所大学的2 883个网页组成的链接网络,图2中3种不同颜色的节点对应了不同大学的网络。

参考文献[12]从社团检测的角度设计网络表示学习算法,目的在于希望学习到的网络表示向量的每一维度都代表一个社团所占的权重。其目标函数是希望模块度最大化<sup>[13]</sup>,最终将问题转化为选取模块度矩阵的前 $k$ 个特征值对应的特征向量作为网络特征表示。

### 3.2 基于最优化的网络表示学习

基于最优化的网络表示学习算法是指根据一个明确的优化目标函数,并且以节点在低维空间的向量表示作为参数,通过求解目标函数的最大化或最小化,求出节点的低维表示的一类算法。下面介绍两个代表性的工作。

网络表示学习的一个重要应用是节点标签预测<sup>[14]</sup>。通常的情景是,在一个网络上每个节点都有对应的标签,但是只有部分节点的标签是已知的,其余节点的标签是未知的,针对这个任务,参考文献[15]中提出了LSHM(latent space heterogeneous model)算法。LSHM算法同时学习节点的向量表示和标签的线性分类函数,它的优化目标函数包括两部分:一部分考虑了网络上的平滑性,即相邻节点的标签尽可能相似;另一部分考虑了分类函数对已知标签的预测能力,因此LSHM算法

是一种半监督的网络表示学习算法。如果只考虑目标函数的第一部分,那么LSHM算法就可以以相似度矩阵作为输入。

LSHM算法可以处理异构网络,图3给出了一个异构网络的示例,这个异构网络包括作者、论文和词语3种类型的节点,作者都有一个表示其所属研究领域的标签,论文都有一个会议标签。LSHM算法处理异构网络的思想是无论节点属于何种类型,都对节点在同一向量空间学习低维表示。LSHM算法提出了一种针对节点属性的扩展方式,这里对节点属性和节点标签进行区分,比如引文网络中,文章的文本内容是节点属性,文章所属的会议是节点标签。LSHM算法针对节点属性的扩展方式是把属性也作为一种没有标签的节点,比如文本属性中的每一个词语都对应一个新的节点,图3中的关键词节点即论文节点的属性,文章节点与词语节点的权重可以定义为词频等指标,这样新的网络依然可以用LSHM算法处理。

网络信息传播预测中,一类传统的方法是首先从用户的行为中发现传播的隐式结构<sup>[16]</sup>,直观上说就是建立一个网络,然后在网络上模拟信息扩散的过程。参考文献[17]学习用户节点在连续隐空间上的低维表示,把在网络上的扩散问题转化为在隐空间上的扩散问题。如图4所示,每个用户都对应隐空间中的一个向量(也可以称作位置、点)。

以上两个基于最优化的网络表示学习算法,各自和特定的网络分析任务有关,在设计优化目标函数时都使用hinge损失函数,求解方法通常是随机梯度下降。

### 3.3 基于概率生成式模型的网络表示学习

概率生成式算法是指用一个基于概率

的生成过程去建模观测数据的产生过程,经典的PLSA、LDA都属于概率生成式算法,本质上都是概率图模型。基于概率生成式算法的网络表示学习,是指用一个采样过程去建模网络数据的生成过程。这些模型的求解方法通常是Gibbs 采样、变分推断和期望最大化算法等。

考虑到概率生成模型通常被用来建模文本数据,基于概率生成式算法的网络表示学习通常也是以文本网络作为输入数据,文本网络是指网络中的每个节点都有对应的文本属性,典型的例子如:网页超

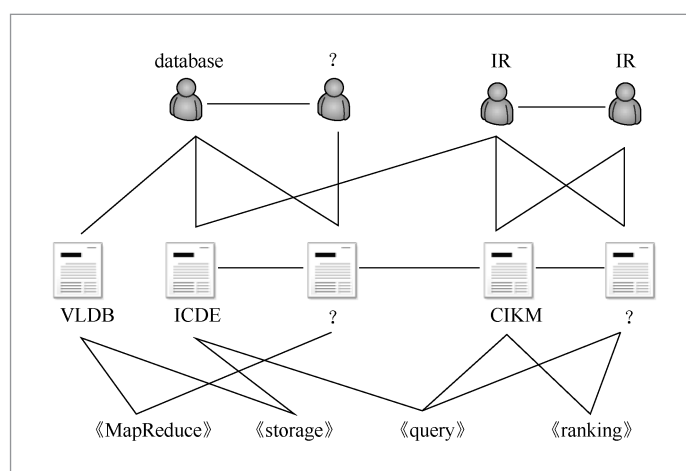


图3 异构网络示例<sup>[15]</sup>

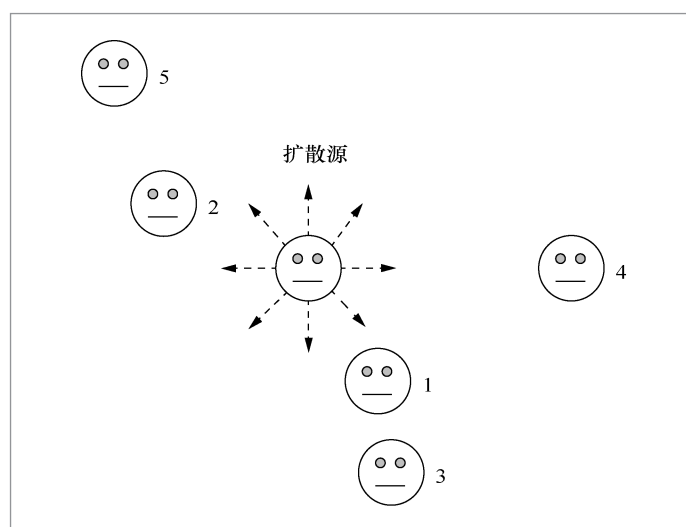


图4 隐空间上的信息扩散<sup>[17]</sup>

链接网络中的每个网页都有自己的文本内容, 在线社会中每个用户都有自己的文本属性。当前针对文本网络的表示学习, 通常学习到的都是文本节点在主题空间上的向量表示, 下面介绍3个代表性的工作: Link-PLSA-LDA模型<sup>[18]</sup>、RTM模型<sup>[19]</sup>和PLANE模型<sup>[9]</sup>。

Link-PLSA-LDA模型是为学术论文引用网络设计的, 其图模型如图5所示, 这

个模型首先建模被引用论文集合的生成过程, 然后建模引用论文集合的过程。

RTM模型建模文本的生成过程与传统的主题模型LDA保持一致, 不同之处在于RTM建模了链接关系的产生, 这个模型的假设是如果两个文本节点之间存在边, 那么它们在主题上的分布应该更相似。PLANE和RTM的图模型表示的对比如图6所示。

PLANE模型对RTM模型进行了扩展, 希望从可视化的角度学习主题和文本节点的低维表示。PLANE模型的生成过程希望同时建模两个方面: 一是表示学习, 即对每个主题和每个文本节点都学习其在二维空间上的坐标, 二是传统的主题建模。RTM模型不能直接得到文本节点在二维空间上的表示, 故PLANE的作者用PE算法<sup>[20]</sup>对RTM模型学习到的文章话题表示进行降维。

如果只有网络的结构信息, 而没有节点各自的文本内容, 那么上面提到的3种算法都无法直接使用, 因为这些算法需要学习节点在主题空间上的分布。解决这个问题

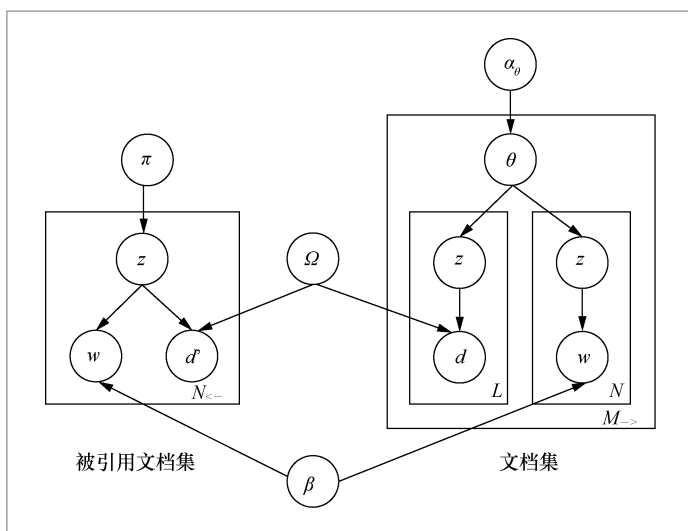


图5 Link-PLSA-LDA的图模型<sup>[18]</sup>

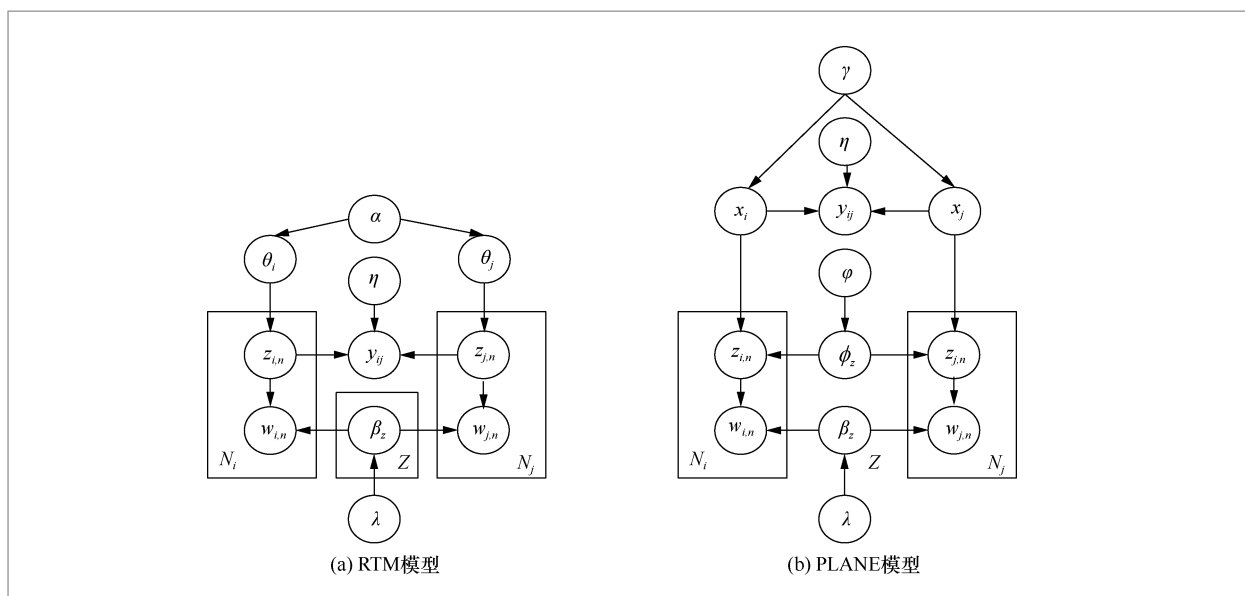


图6 RTM和PLANE的图模型对比<sup>[9]</sup>

题的一种思路是学习节点在社团层面上的分布<sup>[21, 22]</sup>, 这类算法假设每个节点在社团空间上有一个分布, 然后建模网络中边的产生。参考文献[23]进一步考虑了文本内容, 在学习节点在社团空间上的分布的同时, 可以学习社团在主题空间上的分布。

### 3.4 基于力导向绘图的网络表示学习

力导向绘图(force-directed graph drawing)是指一类在美学上让人感到舒适的节点绘图方法, 通常是基于网络中节点的相对位置, 在节点之间和边之间分配作用力, 从而将网络中的节点以某种方式放置在二维或者三维空间中, 使得边的长度尽量相等, 同时尽可能减少交叉边的数量<sup>[24]</sup>。

代表性的方法有FR-layout<sup>[25]</sup>和KK-layout<sup>[26]</sup>, 图7给出了这两种算法对同一网络的可视化效果。这些方法将网络视为一个弹簧系统, 对节点位置进行调整的最终目标是希望这个系统的能量最小化。此类方法通常会被应用到网络可视化软件中<sup>[27, 28]</sup>, 而与机器学习、数据挖掘等任务没有直接的应用关系。

### 3.5 基于深度学习的网络表示学习

近年来, 深度学习技术在语音处理、

图像识别、自然语音处理等领域掀起了巨大的热潮。深度学习本质上是一种特征学习方法<sup>[29]</sup>, 其思想在于将原始数据通过非线性模型转变为更高层次的特征表示, 从而获得更抽象的表达。与特征工程中需要人工设计特征不同, 深度学习会自动从数据中学习出特征表示。

在自然语言处理领域, 一个重要的研究方向是学习词语、句子、文章等的分布式向量表示。针对网络表示学习, 2014年以来也出现了两个具有代表性的基于深度学习的模型: Deepwalk<sup>[30]</sup>、LINE<sup>[31]</sup>。这两个模型都是基于当前最流行的神经网络语言模型word2vec<sup>[32-34]</sup>, 下面首先简要回顾神经网络语言模型的发展过程, 然后介绍神经网络语言模型在网络表示学习中的应用。

学习分布式表示的思想最早来源于Hinton在1986年的工作<sup>[35]</sup>, 利用神经网络语言模型学习词语的词向量则是Bengio在2001年提出的<sup>[36]</sup>。图8是Bengio等人设计的一个4层神经网络语言模型(NNLM)。这个模型用一个长度为 $n$ (这里 $n=4$ )的窗口在语料中滑动, 然后用前 $n-1$ 个词预测观测到窗口内的最后一个词的概率。输入层是窗口内前 $n-1$ 个词的向量表示, 投影层将输入层的向量进行拼接, 隐藏层对投影层的输出进行了非线性转换, 输出层是每个词出现在下一个位置的概率。这个模型

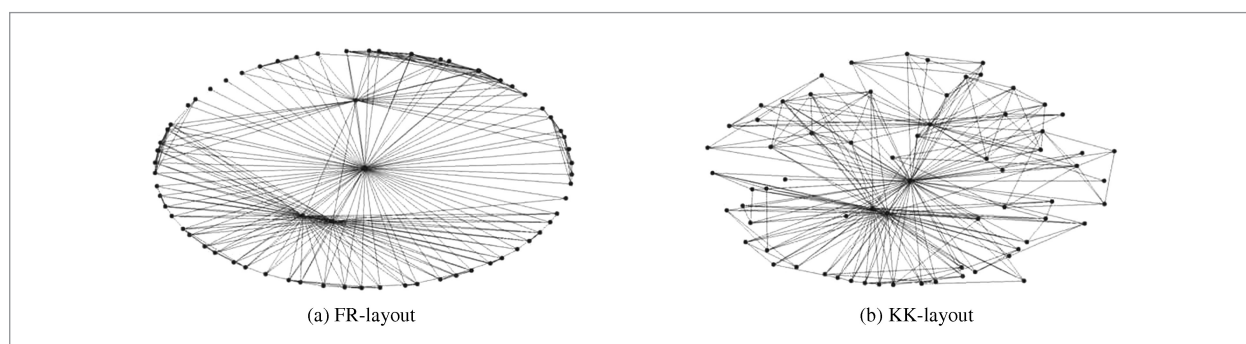


图7 FR-layout 与 KK-layout 示意

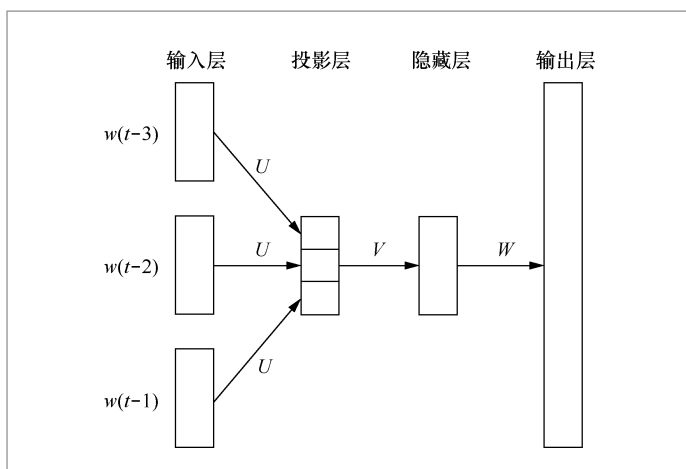


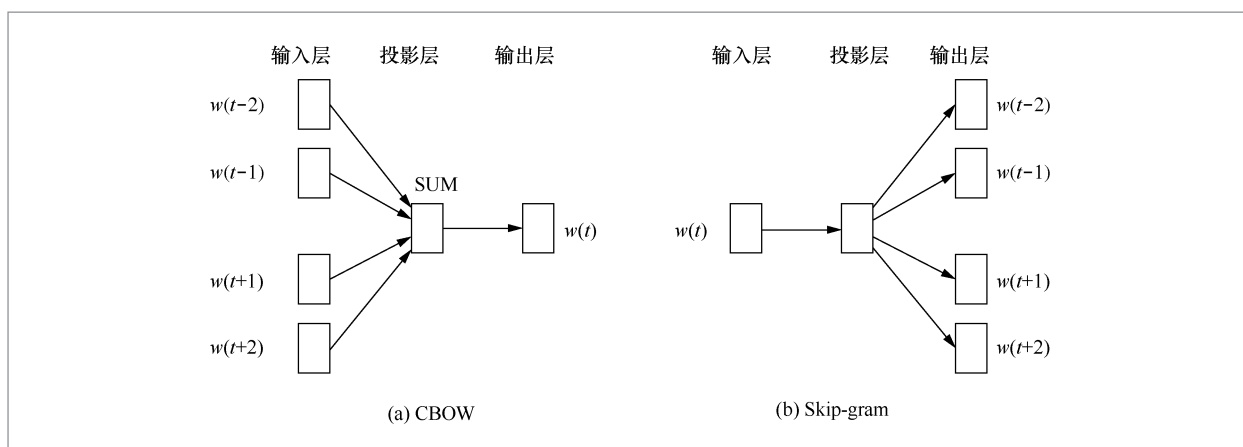
图8 Bengio 提出的4层神经网络语言模型

的求解方法用的是随机梯度下降和反向传播,其中词向量仅是这个神经网络语言模型的副产品。

Bengio提出的这种前向神经网络语言模型的复杂度很高,特别是输出层Softmax计算的复杂度是和词汇表大小同阶。在多个后续工作中,都对Softmax计算进行了替换或优化。如参考文献[37]提出的Hierarchical Softmax,这个算法将输出层建模为一棵霍夫曼树。参考文献[38]中使用了hinge loss函数。Noise contrastive estimation<sup>[39]</sup>也是一种常用的降低NNLM时间复杂度的算法。

近年来最受关注的神经网络语言模型是Mikolov提出的word2vec, word2vec去掉了前向神经网络中的隐藏层,使得训练词向量的速度大幅提高。如图9所示, word2vec包括两个不同的模型,一个是CBOW模型(continuous bag-of-words mode),另一个是Skip-gram模型(continuous skip-gram model)。CBOW模型利用窗口中间词的上下文预测中间的词, Skip-gram反其道行之,用窗口中间的词去预测这个词的上下文。

在词向量学习任务中,输入是文本语料,在网络表示学习任务中,输入是一个网络,看上去这两个任务毫不相关,但Deepwalk算法的出现,解决了这两个任务之间的鸿沟。Deepwalk算法的作者观察到在文本语料中词语出现的频率服从幂律分布,而在无标度网络上进行随机游走的话,节点被访问到的次数也服从幂律分布。因此Deepwalk把节点作为一种人造语言的单词,通过在网络中进行随机游走,获得随机游走路径,把节点作为单词,把随机游走路径作为句子,这样获得的数据就可以直接作为word2vec算法的输入以训练节点的向量表示。这是一个非常具有创造性的想法。图10给出了Deepwalk模型的框架示意。

图9 word2vec 的两种模型<sup>[37]</sup>

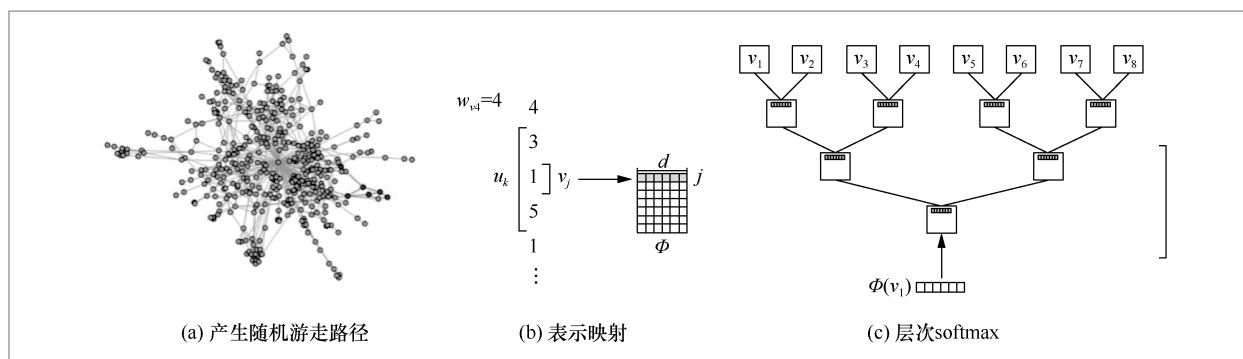


图 10 Deepwalk 模型框架

Deepwalk算法被证明等价于分解矩阵  $M$ ，其中， $M$  的第  $i$  行第  $j$  列的元素是节点  $i$  在固定步数内随机游走到节点  $j$  的次数<sup>[40]</sup>。同样，word2vec也被证明等价于分解PMI矩阵<sup>[41]</sup>或者词共现矩阵<sup>[42]</sup>。在矩阵分解模型框架下，Yang等人提出了在Deepwalk中考虑节点的文本信息的算法TADW模型<sup>[43]</sup>。Deepwalk和TADW的矩阵分解示意如图11所示，其中矩阵  $M$  都可以由网络的邻接矩阵导出。

Deepwalk中矩阵  $M$  被分解为两个矩阵的乘积，这两个矩阵进行拼接后作为节点最终的向量表示，所用的求解方法是正则化的低秩矩阵分解<sup>[44]</sup>。TADW将矩阵  $M$  分

解为3个矩阵的乘积，其中，矩阵  $T$  是节点的文本特征矩阵，通过对TF-IDF矩阵进行奇异值分解降维得到的，计算  $H$  和  $T$  的乘积，把得到的矩阵和  $W$  拼接后得到的矩阵被作为节点的特征矩阵。

Deepwalk在网络中基于随机游走获得训练数据，因此针对网络本身的结构没有一个明确的优化目标函数。参考文献[32]提出的LINE算法，认为网络中存在两种接近度，一种是first-order接近度，是指如果网络中两个节点之间存在边，那么它们之间的first-order接近度是这条边的权重，没有边相连则接近度等于0，这可以看作“物以类聚，人以群分”的一种体现。

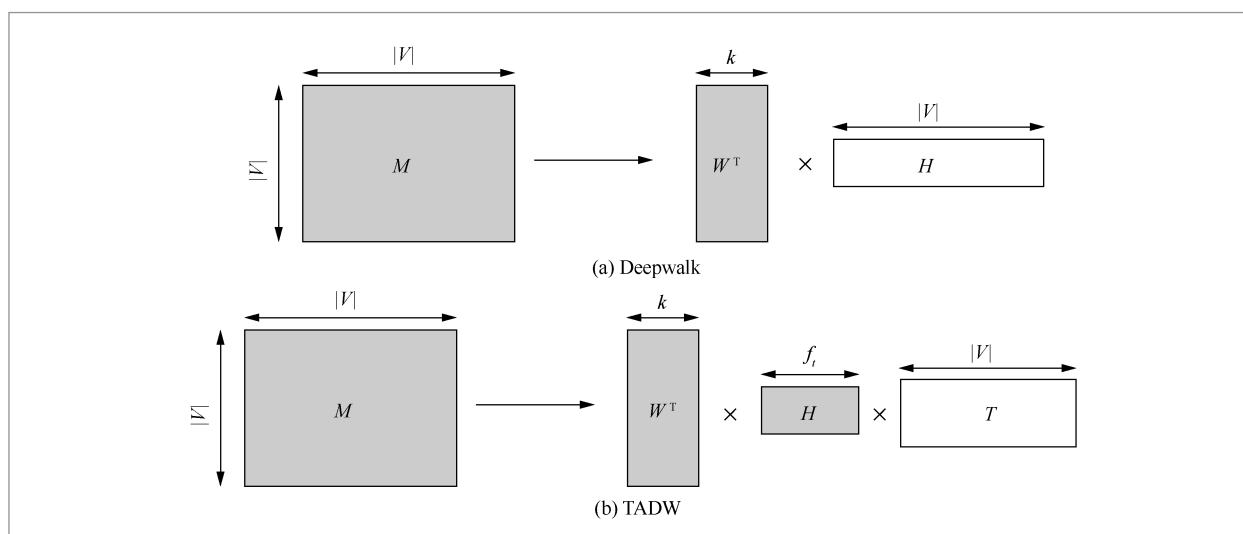


图 11 Deepwalk 和 TADW 的矩阵分解示意<sup>[43]</sup>



就是主题或者社团的个数。在基于深度学习的网络表示学习算法中,  $k$  的取值一般是3位数。

## 5 结束语

本文总结了网络表示学习的主要方法,并特别介绍了基于深度学习的网络表示学习的最新进展。深度学习在社会网络分析领域的应用还处于方兴未艾的探索阶段,但已有工作的结果是令人感到鼓舞的,基于深度学习的网络表示学习算法(如Deepwalk、LINE)在网络节点标签预测任务上的表现,已经超越了传统的基于谱的方法<sup>[13]</sup>、基于最优化的方法<sup>[46]</sup>。然而,这两个算法缺少在其他社会网络分析任务中的应用案例,普适性不足。基于以上分析,本文设想了未来两个值得探索的研究方向。

虽然标签预测任务十分常见,但是SNA领域还有很多其他重要的任务,比如影响力最大化、链接预测、社团检测等,这些任务和标签预测任务有着很大的不同:一方面,可以考虑针对特定的社会网络分析任务,设计网络表示学习模型;另一方面,无监督的网络表示学习算法是否对多种任务具有普适性也值得探究。

现实中存在着大量的异构网络,而且节点都对应着丰富的标签、文本、图像、语音等多媒体信息。如何使基于深度学习的网络表示学习算法可以同时利用网络的结构和节点自身的特征信息,是一个重要的问题。这涉及如何设计网络结构,如何从网络中采样节点对等细节。TADW算法通过和Deepwalk等价的矩阵分解方法,在同构网络的表示学习中融入了节点自身的文本信息,这也为从矩阵分解的视角看待网络表示学习的扩展提供了有益的启示。

## 参考文献

- [1] Mairal J, Ponce J, Sapiro G, *et al.* Supervised dictionary learning. Proceedings of the 2009 Conference on Neural Information Processing Systems, Vancouver, Canada, 2009:1033~1040
- [2] Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 2000, 290(5): 2323~2326
- [3] Hyvärinen A, Oja E. Independent component analysis: algorithms and applications. *Neural Networks*, 2000, 13(4~5): 411~430
- [4] Lee H, Battle A, Rain R, *et al.* Efficient sparse coding algorithms. Proceedings of the 2006 Conference on Neural Information Processing Systems. Vancouver, Canada, 2006:801~808
- [5] Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(8): 1798~1828
- [6] Chen M, Yang Q, Tang X O. Directed graph embedding. Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI), Hyderabad, India, 2007:2707~2712
- [7] Kannan R, Vempala S. Spectral algorithms. *Theoretical Computer Science*, 2009, 4(3~4):157~288
- [8] Brand M, Huang K. A unifying theorem for spectral embedding and clustering. Proceedings of the 9th International Conference on Workshop on Artificial Intelligence and Statistics, Florida, USA, 2003
- [9] Le T, Lauw H W. Probabilistic latent document network embedding. Proceedings of 2014 IEEE International

- Conference on Data Mining (ICDM), Shenzhen, China, 2014:270~279
- [10] Wojciech C, Brooks M J. A note on the locally linear embedding algorithm. *International Journal of Pattern Recognition and Artificial Intelligence*, 2009, 23(8): 1739~1752
- [11] Belkin M, Niyogi P. Laplacian eigenmaps and spectral techniques for embedding and clustering. *Proceedings of Annual Conference on Neural Information Processing Systems(NIPS)*, Cambridge, UK, 2001:585~591
- [12] Tang L, Liu H. Relational learning via latent social dimensions. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, 2009:817~826*
- [13] Newman M. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 2006, 103(23): 8577~8582
- [14] Zhou D Y, Huang J Y, Schölkopf B. Learning from labeled and unlabeled data on a directed graph. *Proceedings of the 22nd International Conference on Machine Learning, Bonn, Germany, 2005: 1036~1043*
- [15] Jacob Y, Denoyer L, Gallinari P. Learning latent representations of nodes for classifying in heterogeneous social networks. *Proceedings of the 7th ACM International Conference on Web Search and Data Mining, New York, USA, 2014: 373~382*
- [16] Yang J, Leskovec J. Modeling information diffusion in implicit networks. *Proceedings of 2010 IEEE 10th International Conference on Data Mining (ICDM)*, Sydney, Australia, 2010: 599~608
- [17] Bourigault S, Lagnier C, Lamprier S, *et al.* Learning social network embeddings for predicting information diffusion. *Proceedings of the 7th ACM International Conference on Web Search and Data Mining, New York, USA, 2014: 393~402*
- [18] Nallapati R, Ahmed A, Xing E, *et al.* Joint latent topic models for text and citations. *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Las Vegas, USA, 2008: 542~550*
- [19] Chang J, Blei D. Relational topic models for document networks. *Proceedings of International Conference on Artificial Intelligence and Statistics, Clearwater Beach, Florida, USA, 2009: 81~88*
- [20] Iwata T, Saito K, Ueda N, *et al.* Parametric embedding for class visualization. *Neural Computation*, 2007, 19(9): 2536~2556
- [21] Gopalan P, Blei D. Efficient discovery of overlapping communities in massive networks. *Proceedings of the National Academy of Sciences*, 2013, 110(36): 14534~14539
- [22] Gopalan P, Mimno D, Gerrish S, *et al.* Scalable inference of overlapping communities. *Proceedings of the 2012 Conference on Neural Information Processing Systems, Lake Tahoe, USA, 2012: 2249~2257*
- [23] Hu Z T, Yao J J, Cui B, *et al.* Community level diffusion extraction. *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, Melbourne, Victoria, Australia, 2015: 1555~1569*
- [24] Kobourov S. Spring embedders and force directed graph drawing algorithms. *arXiv Preprint 2012, arXiv:1201.3011, 2012*
- [25] Fruchterman T, Reingold E. Graph drawing by force-directed placement. *Software-Practice & Experience*, 1991, 21(11): 1129~1164
- [26] Kamada T, Kawai S. An algorithm for drawing general undirected graphs. *Information Processing Letters*, 1989, 31(1): 7~15

- [27] Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. Proceedings of the 3rd International Conference on Weblogs and Social Media, San Jose, California, USA, 2009: 361~362
- [28] Ellson J, Gansner E, Koutsofios L, *et al.* Graphviz—open source graph drawing tools. Graph Drawing. Berlin Heidelberg: Springer, 2002
- [29] Bengio Y, Goodfellow I, Courville A. Deep Learning, 2015
- [30] Perozzi B, Al-Rfou R, Skiena S. Deepwalk: online learning of social representations. Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA, 2014: 701~710
- [31] Tang J, Qu M, Wang M Z, *et al.* LINE: large-scale information network embedding. Proceedings of the 24th International Conference on World Wide Web, Florence, Italy, 2015: 1067~1077
- [32] Mikolov T, Sutskever I, Chen K, *et al.* Distributed representations of words and phrases and their compositionality. Proceedings of the 2013 Conference on Neural Information Processing Systems, Lake Tahoe, USA, 2013: 3111~3119
- [33] Mikolov T, Chen K, Corrado G, *et al.* Efficient estimation of word representations in vector space. arXiv Preprint arXiv:1301.3781, 2013
- [34] Mikolov T, Yih W T, Zweig G. Linguistic regularities in continuous space word representations. Proceedings of the 2013 Conference on NAACL and SEM, Atlanta, USA, 2013: 746~751
- [35] Hinton G E. Learning distributed representations of concepts. Proceedings of the Eighth Annual Conference on the Cognitive Science Society, Amherst, Mass, USA, 1986: 1~12
- [36] Bengio Y, Ducharme R, Vincent P, *et al.* A neural probabilistic language model. Journal of Machine Learning Research, 2003(3): 1137~1155
- [37] Morin F, Bengio Y. Hierarchical probabilistic neural network language model. Proceedings of the 10th International Workshop Conference on Artificial Intelligence and Statistics, Barbados, 2005: 246~252
- [38] Collobert R, Weston J. A unified architecture for natural language processing: deep neural networks with multitask learning. Proceedings of the 25th International Conference on Machine Learning, Helsinki, Finland, 2008: 160~167
- [39] Gutmann M, Hyvärinen A. Noise-contrastive estimation: a new estimation principle for unnormalized statistical models. Proceedings on International Conference on Artificial Intelligence and Statistics, Sardinia, Italy, 2010: 297~304
- [40] Yang C, Liu Z Y. Comprehend deepwalk as matrix factorization. arXiv Preprint arXiv:1501.00358, 2015
- [41] Goldberg Y, Levy O. Word2vec explained: deriving Mikolov et al.'s negative-sampling word-embedding method. arXiv Preprint arXiv:1402.3722, 2014
- [42] Li Y T, Xu L L, Tian F, *et al.* Word embedding revisited: a new representation learning and explicit matrix factorization perspective. Proceedings of the 24th International Joint Conference on Artificial Intelligence, Buenos Aires, Argentina, 2015: 3650~3656
- [43] Yang C, Liu Z Y, Zhao D L, *et al.* Network representation learning with rich text information. Proceedings of the 24th International Joint Conference on Artificial Intelligence, Buenos Aires, Argentina, 2015: 2111~2117
- [44] Yu H F, Jain P, Kar P, *et al.* Large-scale multi-label learning with missing labels. arXiv Preprint arXiv:1307.5101, 2013

[45] Tang J, Qu M, Mei Q Z. PTE: predictive text embedding through large-scale heterogeneous text networks. Proceedings of the 21st ACM SIGKDD Conference on knowledge Discovery and Data Mining, Sydney, Australia, 2015

[46] Ahmed A, Shervashidze N, Narayanamurthy S, *et al.* Distributed large-scale natural graph factorization. Proceedings of the 22nd International Conference on World Wide Web, Rio, Brazil, 2013: 37~48

#### 作者简介



陈维政, 男, 北京大学博士生, 主要研究方向为机器学习和社会网络分析。



张岩, 男, 北京大学教授、博士生导师, 主要研究方向为信息检索、文本分析和数据挖掘。



李晓明, 男, 北京大学教授、博士生导师, 主要研究方向为搜索引擎、网络数据挖掘和并行与分布式系统。

收稿日期: 2015-08-20

基金项目: 国家重点基础研究发展计划(“973”计划)基金资助项目(No.2014CB340400), 国家自然科学基金资助项目(No.61272340, No.61472013, No.61532001)

**Foundations Items:** The National Basic Research Program of China (973 Program)(No.2014CB340400), The National Natural Science Foundation of China (No.61272340, No.61472013, No.61532001)

论文引用格式: 陈维政, 张岩, 李晓明. 网络表示学习. 大数据, 2015025

Chen W Z, Zhang Y, Li X M. Network representation learning. Big Data Research, 2015025