

基于网络特征学习的个性化推荐系统

王鸿伟

摘要

在当今信息爆炸的时代，个性化推荐系统是面向用户的互联网产品的核心技术。推荐系统可以帮助用户获取所需要的信息，改善信息超载的问题。推荐系统的技术核心是对用户历史、物品属性和上下文等信息进行建模，推断出用户的兴趣爱好，并向用户推荐感兴趣的物品。因此，实用的推荐算法需要有很强的扩展性，可以方便地融合各种辅助信息。在众多的辅助信息中，有一类较为特殊，即拥有网络结构的信息，例如，用户之间的在线社交网络，以及物品之间的知识图谱，甚至用户和物品的交互本身就是一个二分图。网络结构的信息为推荐算法提供了丰富的辅助输入，然而如何有效地利用这种高维结构数据，成为推荐系统中的一个富有挑战性的问题。

近年来，网络特征学习逐渐成为机器学习中的一个热门的研究方向。网络特征学习试图为一个网络中的每一个节点学习得到一个低维表示向量，同时保持其原有的结构信息。由于推荐系统中天然存在着大量的网络结构，因此，将网络特征学习与推荐系统相结合，用网络特征学习的方法去处理推荐系统中的相关特征，可以有效地增强推荐系统的学习能力，提高推荐系统的精确度和用户满意度，从而为现实生活中的各类互联网应用提供更优良的用户体验，进而减轻信息爆炸带来的负面影响，提升整体经济效率。

本综述将首先介绍推荐系统的相关研究，然后介绍近年来网络特征学习的进展。最后将对网络特征学习在推荐系统中的应用研究现状进行简要介绍。

关键词：推荐系统（recommender systems）；网络特征学习（network representation learning）；机器学习（machine learning）；深度学习（deep learning）

1 推荐系统

在本节中，我们首先介绍推荐系统的背景知识，然后，我们遵循推荐系统中的一般分类方法对其进行介绍，其中我们会重点关注对协同过滤方法。最后，我们简单地介绍一些推荐系统最新的研究领域和方向。

1.1 背景

随着互联网技术和产业的迅速发展，接入互联网的服务器数量和网页数量也呈指数上

升趋势。互联网的迅速发展使得海量信息同时呈现在我们面前。例如，Netflix¹上有数万部电影，Amazon²上有数百万本图书，淘宝网³上的商品更是以亿计数。传统的搜索算法只能呈现给用户（user）一样的物品（item）排序结果，无法针对不同用户的兴趣爱好提供相应的服务。信息爆炸使得信息的利用率反而降低，这种现象被称为信息超载（information overload）。推荐问题从本质上说就是代替用户评估其从未看过、接触过和使用过的物品，包括书籍、电影、新闻、音乐、餐馆、旅游景点等。**推荐系统（recommender systems）**作为一种信息过滤的重要手段，是当前解决信息超载问题的最有效的方法之一，是面向用户的互联网产品的核心技术。

推荐系统在20世纪90年代中期被作为一个独立的概念提出[1, 2, 3]。二十年以来，推荐系统的研究在学术界和工业界得到迅速发展，诞生了大量的理论、技术和应用。2006-2009年，Netflix推出的推荐系统比赛吸引了超过41000支队伍[4]，极大的提高了推荐系统在学术领域的关注度。推荐系统本质上是一种信息过滤系统，其核心是对用户历史、用户特征、物品属性、上下文信息等进行建模，以推断出用户兴趣，并向用户推荐其感兴趣的物品。在信息大数据的时代，要有效地利用海量的用户历史信息和物品信息，发掘用户和物品之间的交互规律，**机器学习（machine learning）**技术必不可少。相应地，推荐系统作为机器学习的核心应用之一，也极大地促进了机器学习技术的进步。

1.2 传统推荐系统方法分类

在早期的综述中[5]，推荐系统被分为三类：基于内容的方法（content-based methods）、协同过滤（collaborative filtering）、和混合方法（hybrid methods）。近年来，随着可用信息的增加，推荐算法日益细化和复杂，单纯的基于内容的方法或协同过滤越来越少，综合利用各种可用信息是推荐系统的未来趋势。在本文中，我们首先依然遵循传统分类方法，对推荐系统进行介绍。

1.2.1 基于内容的方法

基于内容的推荐系统使用针对物品（或用户）的描述来定义物品（或用户）的属性[6, 7]。推荐系统会根据用户过去给出评分最高的物品的属性来给用户推荐最为匹配的物品。实现这种推荐方法的关键之一在于定义一种相似度度量标准（例如余弦距离）。

基于内容的方法的优势在于对用户兴趣可以很好的建模，并可以通过对物品属性维度的增加，获得更好的推荐精度。不足之处在于：一，物品的属性有限，很难获得更多的属性数据，而对于很多类型的数据而言（图像，视频），自动特征提取是一件比较困难的事情[3]⁴；二，基于内容的方法只会给用户推荐和他的属性高度匹配的物品，也就是说，该推荐方法无法帮助用户发现新的兴趣和爱好。

¹<https://www.netflix.com>

²<https://www.amazon.com>

³<https://www.taobao.com>

⁴这种说法是针对传统推荐算法而言的。近五年来，随着计算能力的不断增强和数据量的空前增长，深度学习已经在针对图像和视频的特征学习领域取得了重大的突破，将图像等曾经难以提取特征的辅助信息用到了推荐系统中也成为了可能。

1.2.2 协同过滤

协同过滤是推荐系统中最为流行和广泛使用的技术。简单地说，协同过滤方法根据其他相似用户对于某物品的评分来预测某用户对于该物品的评分。协同过滤方法基于如下的两个假设：一，在过去拥有相似偏好的用户在未来也有相似的偏好；二，用户对偏好随时间变化相对稳定。早期的综述将协同过滤方法分成两大分类：基于记忆（memory-based）的算法和基于模型（model-based）的算法。

基于记忆的算法

基于记忆的算法[2, 8, 9, 10]根据其它用户对某物品的评分来预测某用户对该物品的评分。也就是说，对于用户 c 和物品 s ，一个未知评分 $r_{c,s}$ 的预测值为其他用户（通常为若干个最相似用户）对于该物品 s 的评分的函数：

$$r_{c,s} = \underset{c' \in C}{aggr} r_{c',s}, \quad (1)$$

其中 C 表示若干个对 s 给出评分的且和 c 最相似的用户集合。聚合函数 $aggr$ 可以是简单的算术平均值，基于相似度的加权平均值等。相似度的度量可以是皮尔逊相关系数，余弦距离等。

基于模型的算法

基于模型的算法通过机器学习方法，使用评分集合学习出一个全局模型，然后基于模型来预测缺失的评分。由于可以选择的机器学习模型千差万别，基于模型的算法也有非常多的种类。[11]提出了一种基于支持向量机（SVM）的方法，使用启发式算法迭代计算估计缺失的评分值。[12]使用一种神经网络的算法，基于用户和基于物品分别进行预测。[13]提出了各种针对Netflix竞赛的推荐系统算法，并特别指出了矩阵分解算法给出了最好的结果。其它在推荐系统中使用的模型还有：贝叶斯模型[14]，概率关系模型[15]，线性回归[16]，最大熵模型[17]，决策模型[18]，概率语言（PLSA）模型[19, 20]，LDA（Latent Dirichlet Allocation）模型[21]等。

协同过滤建模主要使用用户对物品的历史交互数据，也称为反馈数据。根据交互行为是否反映用户对物品的喜好程度，可以把反馈数据分为两类：（1）显式反馈（explicit feedback），通常是指能够直接反映用户对物品喜好程度的评分，例如豆瓣网⁵上用户对电影的1-5分的评分；（2）隐式反馈（implicit feedback），例如点击、购买、看视频、听音乐等行为。这些行为不能直接揭示用户是否喜欢一个物品，但能从侧面反映出用户对物品的兴趣。

显式反馈

早期较为常见的推荐系统任务是评分预测[22]。输入为三元组集合 $\langle u, i, y_{ui} \rangle$ ，代表用户 u 对物品 i 的打分 y_{ui} 。所有历史数据可以表示为一个二维矩阵 Y ，其中行、列和值分别代表用户、物品和打分。由于互联网产品中用户和物品的规模通常很大（百万级甚至更多），而

⁵<https://www.douban.com>

	Item I			
Context C	2		3	
			4	
		5		2
	3	1		4
			3	

	Item I			
Context C	1	0	1	0
	0	0	1	0
	0	2	0	1
	1	1	0	2
	0	0	1	0

图 1: 显式反馈和隐式反馈的区别[27]: 左图为显式反馈, 矩阵中每个元素1-5代表用户对物品的喜好程度; 右图为隐式反馈, 矩阵中每个元素1或0代表用户是否选择了某一物品。

用户通常只对少量的物品有打分, 因此 Y 矩阵通常高度稀疏。评分预测任务可以表示为预测打分矩阵中的缺失数据。

通常而言, 基于显式反馈的评分预测系统只优化 Y 矩阵中的观察数据, 而完全忽略缺失数据。虽然这种做法极大降低了模型训练的时间复杂度并在预测未知评分任务上可以取得较低的误差率, 但是在以排序为主的top-k物品推荐任务上表现较差[23], 甚至弱于非个性化的基于物品流行度的排序[24]。其主要原因是观察数据中有较强的选择偏差 (selection bias), 而缺失数据中含有丰富的负样本信息[25]。因此, 在构造实际的top-k物品推荐系统时, 传统评分预测模型完全忽略缺失数据的做法并不可取, 考虑对缺失数据的建模非常重要。缺失数据的建模在基于隐式反馈的推荐方法中得到了广泛的研究和使用。

隐式反馈

相比于显式反馈, 互联网内容提供商更容易获得隐式反馈, 例如电商/视频网站可以从服务器中直接获取用户的点击/观看历史。由于不需要用户显式地提供打分, 隐式反馈的选择偏差较小, 而且其规模相对较大。因此, 近些年对推荐系统算法的研究更集中在隐式反馈[26, 27, 28, 29, 30]。

与显式反馈类似, 我们可以将隐式反馈描述为一个二维矩阵 Y 。不同的是这里 Y 中的每一个元素不是一个具体的打分, 而是代表用户是否选择某一物品: 1代表选择, 0代表没有选择⁶。因此, 对隐式反馈的建模更类似于一个二分类问题——预测用户选择某一物品的概率。图1描述了显式反馈和隐式反馈数据上的区别。

值得一提的是, 许多基于显式反馈的预测模型, 如SVD++[22], timeSVD[32]等, 对于隐式反馈同样适用, 前提是要调整其优化目标函数, 以适当的方式将缺失数据考虑进来。

综上所述, 协同过滤方法的优势在于它可以很好地支持用户发现潜在的兴趣偏好。协同过滤的不足之处在于: 一, 对新加入的用户和物品都无法给出准确的推荐, 即冷启动问题 (cold start problem); 二, 在显式反馈中, 由于评分矩阵的稀疏性 (sparsity), 计算结果不是非常稳定, 受异常值影响较大; 三, 仅仅使用评分信息, 没有包含用户和物品本身的属性, 也不包含社交关系信息 (如果可用的话)。

⁶在隐式反馈中, Y 的每个元素也可以代表选择次数, 例如一个用户可能会观看同一个视频多次。相关研究可以参考[31]。

1.2.3 混合方法

鉴于基于内容的方法和协同过滤都存在各自的缺点，一些推荐系统将两者结合，采用混合方法来避免两者的缺陷。不同的结合方法可以将混合推荐系统划分为四种类型：

一，分别实现基于内容的方法和协同过滤，然后结合两者的预测结果。可以使用各个评分预测的线性组合[33] 或设计投票机制[34] 来结合所有的预测结果。

二，在协同过滤中加入一些基于内容的特征。例如，[35] 不仅实现了传统的协同过滤方法，而且维护了每个用户的基于内容的属性。

三，在基于内容的方法中加入一些协同过滤的特征。此类别中最著名的方法是在一系列基于内容的属性中使用降维（dimensionality reduction）技术[36]。

四，构造一个包含基于内容方法和协同过滤的统一模型。例如，[37] 提出在一个基于规则的分类器中结合两者的特征；[38] 和[39] 提出了一个结合两者的统一概率模型。

1.3 推荐系统最新的研究热点和方向

随着大数据和人工智能热潮的蓬勃发展，推荐系统领域也相应地诞生出一些最新的研究热点，其中包括上下文感知（context-aware）的推荐、跨域（cross-domain）推荐、基于深度学习（deep learning based）的推荐等。

1.3.1 上下文感知的推荐

上下文信息（contextual information）在不同的文献中有不同的定义。文献[40]将上下文定义为研究对象的附近个体的位置、身份等信息。文献[41]将上下文定义为任何可以被用来描述个体所处环境的特征的信息。用户的上下文信息包含不同的属性，例如地理位置、情感状态、生理状态、行为特征、个人历史等[42]。传统推荐系统的关注点在于根据重要度和关联度来推荐给用户合适的物品，而没有考虑到额外的上下文信息。然而在很多推荐场景中，例如旅游景点推荐、移动应用推荐等，只考虑到用户和物品的属性是不够的，将上下文信息包含进推荐系统的工作流程非常重要。

学术界已经有不少文献开始关注上下文感知的推荐系统。[43]中的模型考虑了在人机交互过程中动态的上下文因素对推荐效果的提升。[44]使用了多维张量（用户-物品-上下文）来对用户偏好进行建模。[45]指出，智能手机中个人数字助理（如Siri, Cortana）的目标是在正确的时间提供正确的信息，然而现实中用户的意图高度依赖于上下文，如时空信息、用户当前活动等。[38]提出了一种用户上下文感知推荐的协同临近预报模型。

1.3.2 跨域推荐

传统推荐系统关注于单个平台或领域内的推荐，而使用多个平台或领域上所有可用的用户偏好信息可能会促成更有包容性的用户建模和更准确的推荐，例如，解决目标领域的冷启动问题或稀疏性问题。因此，跨域推荐系统的目标是利用源领域的知识（多数为用户偏好）辅助在目标领域中提高个性化推荐的准确度。

从用户角度而言，跨域推荐大致可以分为三类：一，源领域和目标领域的用户完全重叠；二，源领域和目标领域的用户部分重叠；三，源领域和目标领域的用户完全不重叠。每

种类别都有各自相应的跨域推荐建模方法，读者可以参考综述[46]获得更多信息。我们这里选择几篇有代表性的跨域推荐模型进行介绍。[47]将社交网络表示为多个关系领域，并通过迁移其它领域的知识到个体领域中来减轻稀疏性和冷启动的问题。[48]认为不同的平台不仅共享一个公有的评分模式，每个平台也都有自己特定的模式。因此，[48]提出了一个基于聚类的潜在变量模型来学习两种不同的模式。[49]研究了跨设备的搜索，提出了预测跨设备的搜索迁移的模型。[50]将社交网络表示成一个星形混合图，并从辅助领域将知识迁移到目标领域以提高推荐准确度。

1.3.3 基于深度学习的推荐

深度学习一般指代多层人工神经网络，近年来在语音识别、计算机视觉和自然语言处理等领域取得了巨大成功[51]。根据深度学习在推荐系统中的应用方式，我们可以将相关工作大致分为两类：

一，作为一种基于数据的特征学习方法，深度学习技术可以从语义较为丰富的输入数据（如语音、图片、文本）中抽取也有有效的特征表示，以方便在推荐算法中使用这些辅助信息[52, 53, 54, 55, 56]。例如，[52]提出了利用深度神经网络模型来处理音频并进行音乐推荐；[53]利用自编码器（autoencoder）来得到文本的高层特征，并结合协同过滤框架来进行文章推荐；[55]利用卷积神经网络（convolutional neural networks）对图片进行特征学习，并应用到社交网络上的图片推荐系统中；[56]对于用户和物品的id特征都使用自编码器进行特征学习，同时在自编码器的每一层都加入辅助信息（side information），并对最终的低维特征使用协同过滤进行推荐建模。

二，作为一种通用的数据建模方法，深度学习对数据进行多重非线性变换，可以拟合出较为复杂的预测函数。推荐系统中的核心算法是协同过滤，其目标从机器学习的角度可以看成拟合用户和物品之间的交互函数，因此，近期一系列的工作也将深度学习技术应用于协同过滤的交互函数上[57, 58, 59, 60, 61, 62]。例如，DeepFM[57]扩展了解析机（factorization machine）方法，在FM中引入了一个深度模型来拟合特征之间复杂的交互关系；Wide&Deep[58]的Wide部分采用和解析机一样的线性回归模型，Deep部分采用了基于特征表示学习的多层感知机模型；[59]提出了一个深度模型用于Youtube的视频推荐；NCF[60]用多层感知机替换传统协同过滤中的内积操作；DMF[62]类似于DeepFM，在传统矩阵分解模型中引入了一个深度学习模块来提高模型的表达能力。

2 网络特征学习

在本节中，我们首先介绍网络特征学习的背景知识，然后按照不同的分类方法，详细介绍各类网络特征学习方法的特点。

2.1 背景

网络（network）或图（graph）结构广泛地存在于多种真实场景中，如在线用户的社交网络、学术文章的引用网络、知识图谱等。对网络结构的分析可以让我们更好地挖掘网

络中隐藏的信息，在连接预测（link prediction）、节点分类（node classification）、聚类（clustering）、推荐（recommendation）、可视化（visualization）应用中都有极其广泛的用途。

在原始的网络中，每个节点的邻接向量是一个非常冗长且稀疏的表示，这种特征难以被现有的机器学习方法有效地利用。网络特征学习（network representation learning，也叫作network embedding）试图为一个网络中的每个节点（vertex）学习得到一个低维表示向量（low-dimensional representation vector），同时保持其原有的网络结构信息。由于低维向量可以很容易地被各种机器学习方法处理（例如逻辑斯蒂回归、支持向量机等），因此，网络特征学习近年来逐渐成为一个热门的研究领域。

2000年以前

在2000年以前，网络特征学习的主要表现形式是对高维数据进行降维。传统的方法包括主成分分析（Principle Component Analysis）[63]，线性判别分析（Linear Discriminant Analysis）[64]，多维缩放（Multiple Dimensional Scaling）[65]等。主成分分析[63]经常用于减少数据集的维数，同时保持数据集中的对方差贡献最大的特征。这是通过保留低阶主成分，忽略高阶主成分做到的。这样低阶成分往往能够保留住数据的最重要方面。技术上来说，主成分分析对样本 X 的协方差矩阵 XX^T 进行特征值分解，取最大的 k 个的特征值对应的特征向量，将其作为投影矩阵。主成分分析是一种无监督学习方法，出发点是尽可能保留数据的方差，这对于某些分类问题并不是非常有效。鉴于这一点，线性判别分析[64]的核心思想是将高维空间中的数据点映射到低维空间中，使得同类点之间的距离尽可能接近，不同类点之间的距离尽可能远。线性判别分析的支持向量机有异曲同工之处，区别在于支持向量机是将数据从低维向高维进行映射，目的是为了更好的分类；而线性判别分析是将数据从高维向低维进行映射，目的是为了更好的降维。多维缩放[65]的思想更简单直接，即在低维空间中（近似）保持每一对点的距离等于它们在原始高维空间里的距离。

2000年左右

2000年前后也涌现了一些新的降维方法，其中较为经典的有同态映射（isometric mapping）[66]、局部线性嵌入（locally linear embedding）[67]和拉普拉斯特征映射（Laplacian eigenmaps）[68]。IsoMap和LLE都是流形学习（manifold learning）中的方法。流形学习假设高维数据分布在一个特定的低维空间结构（流形）上，然后试图在低维空间上保持原有高维空间中数据的结构特征。具体来说，IsoMap[66]只考虑高维空间中每个点和它最邻近的 k 个点的距离，基于这些（测地线）距离计算所有点对的距离，然后调用多维缩放的算法进行降维，保持每个节点和其局部邻近节点之间的距离关系。而LLE[67]在高维空间中计算每个点和它邻近节点的线性依赖关系，并试图在低维空间中保持这种线性依赖。拉普拉斯特征映射[68]则是另一大类方法，即图拉普拉斯（graph Laplacian）或谱分析（spectral analysis）的代表。事实上，图拉普拉斯方法在近年来最新的网络特征学习方法中依然占据一席之地。

2010年以后

2010年以后，研究者们对网络特学习的研究也从由高维数据构建起来的“网络”逐渐转移

到真实的网络结构上来。2013年, word2vec[69]提出利用SkipGram模型来学习词的低维向量表示(也叫词向量, word embedding)。由于word2vec简洁的方法和优秀的性能, 学术界掀起了一股特征学习的热潮, 网络特征学习的研究也呈现井喷之势。我们接下来将按照网络特征学习的输入、输出和方法, 对该领域进行简要介绍。

2.2 根据输入分类

网络特征学习的输入是一个网络, 输入网络的类型可以分为如下几类:

一, 同构图, 即网络中的节点和边的种类都只有一个, 例如引用网络。其中同构图又可以进一步按照无向/有向、无权重/有权重、无符号/有符号等标准细分。例如, [70, 71, 72]将图中所有的节点和边都视为同一类别, 整个图中只有基本的结构信息是可用的。[73, 74]考虑的是有权重的图, 即图中的边有附加一个权重信息。两个点之间的权重越大, 它们的表示向量就会越近。[75]考虑的社交网络结构中, 用户之间有关联/被关注的关系, 所以该图是一个有向图。另外, 一些通用的网络特征学习算法会同时适用于无向/有向、无权重/有权重的图, 例如[76, 77, 78]。还有一些工作关注于有符号的网络[79, 80], 这些网络一般是基于情感(sentiment)的数据集, 如购物网站上的评论记录等。

二, 异构图, 即网络中的节点或边的种类不止一个。一般而言, 这种图有两个表现形式: (1) 多媒体网络(multimedia network)。例如, [81]考虑了有两种类别的节点(图像和文本)和三类别的边(图像-图像、图像-文本、文本-文本)的图; [55]考虑了一个由用户和图像组成的网络。(2) 知识图谱(knowledge graph), 知识图谱中, 每个点是一个实体(entity), 每条边表示一种关系(relation), 每个三元组 (h, r, t) 表示头结点 h 和尾节点 t 由关系 r 相连接。由于知识图谱表示学习(knowledge graph embedding)具有重要的应用价值, 近几年来也成为一个热门研究方向, 读者可以参考最新的综述[82]。值得一提的是, 在知识图谱特征学习中, 有一类基于翻译(translation-based)的方法引人注目, 例如TransE[83]、TransH[84]、TransR[85]等, 这类方法的目标是, 针对一个三元组 (h, r, t) , 学习得到的向量满足

$$\mathbf{h} + \mathbf{r} \approx \mathbf{t}, \quad (2)$$

即关系 r 充当了从头结点到尾节点的翻译转换的角色, 因此得名。

三, 带辅助信息的图, 比如节点或边可能会携带标签信息、属性信息、特征信息等。例如, [86]假定节点有一个二分类的标签, 并将网络特征学习和训练分类器在目标函数的层面联合起来。[87]考虑了图中节点和边既有离散的属性又有连续的属性的情况。[88, 89]分别考虑了节点携带了文本特征和文档特征情况下的网络特征学习。

四, 由非关系型数据转化成的图, 这一类常见于早期方法, 即高维数据的降维算法, 详见2.1节。

2.3 根据输出分类

一, 输出节点的特征。网络特征学习中最常见的输出是节点的特征向量。几乎所有前文所述的方法都是此类。

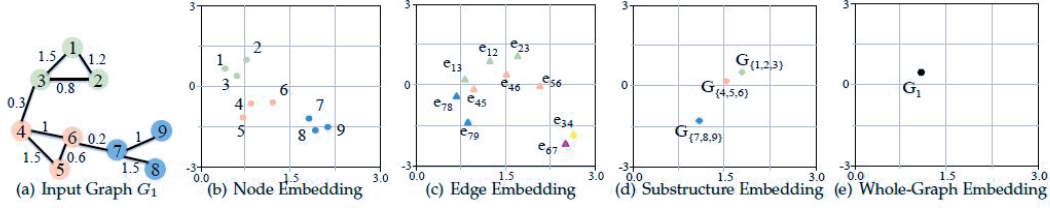


图 2: 将一个网络通过特征学习得到不同粒度的特征表示的示意图。 $G_{\{1,2,3\}}$ 表示包含 v_1 , v_2 和 v_3 的子结构。从左到右分别为: (a) 原图; (b) 节点的特征; (c) 边的特征; (d) 子结构的特征; (e) 全图的特征。

二，输出边的特征。在知识图谱特征学习中，除了输出节点的特征学习之外，也会输出边的特征向量。这是因为知识图谱中，每一条边都属于一种特定的关系（relation），关系的特征向量在知识图谱的各种应用中，例如关系抽取（relation extraction）、知识图谱补全（knowledge graph completion）、问答系统（question answering）中都有重要的作用。另外，如果网络特征学习用于连接预测（link prediction）的任务，那么也会输出一个节点对（即一条边）的特征向量[78]。

三，输出混合特征。这种情况下，方法会输出不同的组成部分的特征组合，例如节点+边（即图的子结构），或节点+团体（community）等。例如，[90]学习了子图（subgraph）的特征，用来为图的分类问题定义一种图内核（graph kernels）。[70]考虑了团体感知（community-aware）的高阶节点近邻关系，并在一个统一的框架中联合优化了团体检测目标和网络特征学习的任务。

四，输出全图的特征。这种情况一般为输出一些小型网络结构（例如蛋白质、分子等）的特征，每一个小型网络结构被表示成单个向量，两个相似的图会得到相似的特征向量。全图特征学习可以直接用来计算图的相似性[91, 87]。

图2描述了不同输出的实例。

2.4 根据方法分类

网络特征学习按照方法可以大致分成如下几类：

一，矩阵分解（matrix factorization）。这种方法涵盖了几乎所有的传统方法，包括主成分分析（PCA）、多维缩放（MDS）、同态映射（IsoMap）等。另外，以拉普拉斯特征映射（Laplacian eigenmaps）为代表谱分析（spectral analysis）方法也是矩阵分解的一种，详见2.1节。

二，随机游走（random walk）。这种方法使用随机游走的方法，从图中采样出很多条路径，然后基于这些路径来学习节点的特征表示。事实上，在将图表示为路径的集合之后，整个图就被转换成了一个有节点组成的“文档”，因此，以word2vec[69]为代表的词向量学习方法都可以被应用到这里。第一个使用随机游走来进行网络特征学习的方法是DeepWalk[92]，DeepWalk正是利用随机游走，将图表示为游走路径的集合，然后使用SkipGram方法进行节点特征学习。Node2vec[78]也是这种方法的代表之一。Node2vec很类似于DeepWalk，区别在于node2vec更细粒度地定义了随机游走中选择下一个节点的方法。

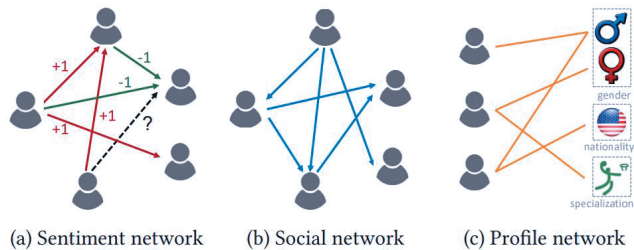


图 3: 预测微博用户对名人情感所使用的三个网络结构示意图。从左到右分别为: (a) 情感网络; (b) 社交网络; (c) 属性网络。

式。

三, 深度学习 (deep learning) 。[93, 94]使用自编码器 (autoencoder) 进行网络特征学习。例如, SDNE[94]将网络中每个节点用它的邻接向量表示, 当做自编码器的输入, 然后对于每一条边, 都让其两个点在自编码器的中间层输出尽量接近。这种方法的一个不足之处在于, 对于大规模网络而言, 邻接向量会非常冗长且稀疏, 从而使这种方法的计算效率大大降低。HNE[81]也利用卷积神经网络 (convolutional neural networks) 来处理图像, 进行网络特征学习。

四, 其它自定义损失函数 (self-defined loss function) 。例如, LINE[77]中定义了一阶邻近关系和二阶邻近关系, 并以此为目标设计了两个损失函数项。知识图谱特征学习中的各种方法也都是自定义的损失函数, 详见2.2节。

3 网络特征学习在推荐系统中的应用现状

网络特征学习在推荐系统中的应用较为少见。在[95]中, 作者考虑了知识图谱中的三种辅助信息: 1.知识实体之间的联系; 2.物品相关的文本特征; 3.物品相关的图像特征。其中, 第一种信息使用TransR[85]方法来提取实体特征, 第二种和第三种信息分别使用去噪自编码器 (stacked denoising autoencoder) 和反卷积自编码器 (stacked deconvolutional autoencoder) 来提取文本和图像特征。这三种特征最后在一个统一的协同过滤的框架下被融入到推荐系统算法中。

接下来, 我们简要介绍一下我们做的一些相关工作。SHINE[96]考虑的是在微博上预测用户对于名人的情感 (sentiment prediction), 并据此来给用户推荐他可能会喜欢的名人。为此, 我们考虑了三种网络结构: (1) 情感网络, 这是一个用户和名人之间的二分图, 其中边是有正负的; (2) 用户端的社交网络; (3) 名人端的属性网络 (即一个简化版的知识图谱)。三种网络的示意图在图3给出。针对这三种不同的网络, 我们设计了一个统一的有符号异构网络特征学习方法 (signed heterogeneous information network embedding), 来联合预测用户对于名人的情感。针对每个网络, 我们都分别设计了自编码器来学习对应的特征, 然后将三者用特定的聚合函数融合起来得到最终的预测情感值。

我们也考虑了直接使用网络特征学习的方法来建模用户和物品直接的交互。在[97]中, 我们受对抗生成网络 (generative adversarial nets) 的启发, 将对抗训练的思想引入了网络

特征学习中。在GraphGAN框架中，生成器用于生成（选择）和某个节点可能会有边存在的点，而判别器用于判断两个点之间是否有边存在。在生成器和判别器的对抗下，生成器会逐渐学习到节点之间隐含的连接性分布，而判别器会为生成器提供相应的监督信号。最终学习得到的节点特征可以直接用于推荐系统的应用。

4 总结

本综述的目标是回顾推荐系统和网络特征学习的相关文献，分别梳理各自的历史、发展脉络和最新研究方向，并揭示两者之间深层的联系。将推荐系统和网络特征学习相结合，利用网络特征学习的技术更好地融合网络结构的辅助信息，可以更好地提高推荐的效率。推荐场景本身包含了丰富的网络结构，例如用户端的社交网络，物品端的知识图谱，以及用户和物品本身的交互就构成了一个二分图结构。如果能够借助一些最近的网络特征学习的方法，并将其有效且自然地融合进推荐算法的框架，那么将很大程度上提升推荐的精确度和用户满意度。综上所述，推荐系统和网络特征学习的结合将会是一个充满前景和挑战的新的研究方向。

参考文献

- [1] W. Hill, L. Stead, M. Rosenstein, and G. Furnas, “Recommending and evaluating choices in a virtual community of use,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 194–201, ACM Press/Addison-Wesley Publishing Co., 1995.
- [2] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, “GroupLens: an open architecture for collaborative filtering of netnews,” in *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, pp. 175–186, ACM, 1994.
- [3] U. Shardanand and P. Maes, “Social information filtering: algorithms for automating “word of mouth”,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 210–217, ACM Press/Addison-Wesley Publishing Co., 1995.
- [4] J. Bennett, S. Lanning, *et al.*, “The netflix prize,” in *Proceedings of KDD cup and workshop*, vol. 2007, p. 35, New York, NY, USA, 2007.
- [5] G. Adomavicius and A. Tuzhilin, “Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions,” *IEEE transactions on knowledge and data engineering*, vol. 17, no. 6, pp. 734–749, 2005.
- [6] P. Lops, M. De Gemmis, and G. Semeraro, “Content-based recommender systems: State of the art and trends,” in *Recommender systems handbook*, pp. 73–105, Springer, 2011.
- [7] M. J. Pazzani and D. Billsus, “Content-based recommendation systems,” in *The adaptive web*, pp. 325–341, Springer, 2007.
- [8] J. S. Breese, D. Heckerman, and C. Kadie, “Empirical analysis of predictive algorithms for collaborative filtering,” in *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, pp. 43–52, Morgan Kaufmann Publishers Inc., 1998.
- [9] J. Delgado and N. Ishii, “Memory-based weighted majority prediction,” in *SIGIR Workshop Recomm. Syst. Citeseer*, 1999.
- [10] A. Nakamura and N. Abe, “Collaborative filtering using weighted majority prediction algorithms,” in *ICML*, vol. 98, pp. 395–403, 1998.

- [11] Z. Xia, Y. Dong, and G. Xing, "Support vector machines for collaborative filtering," in *Proceedings of the 44th annual Southeast regional conference*, pp. 169–174, ACM, 2006.
- [12] M. Kim, E. Kim, and J. Ryu, "A collaborative recommendation based on neural networks," in *Database Systems for Advanced Applications*, pp. 425–430, Springer, 2004.
- [13] J. Kim and H. Park, "Fast nonnegative matrix factorization: An active-set-like method and comparisons," *SIAM Journal on Scientific Computing*, vol. 33, no. 6, pp. 3261–3281, 2011.
- [14] Y.-H. Chien and E. I. George, "A bayesian model for collaborative filtering,," in *AISTATS*, 1999.
- [15] L. Getoor, M. Sahami, *et al.*, "Using probabilistic relational models for collaborative filtering," in *Workshop on Web Usage Analysis and User Profiling (WEBKDD'99)*, 1999.
- [16] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in *Proceedings of the 10th international conference on World Wide Web*, pp. 285–295, ACM, 2001.
- [17] D. Y. Pavlov and D. M. Pennock, "A maximum entropy approach to collaborative filtering in dynamic, sparse, high-dimensional domains," in *Advances in neural information processing systems*, pp. 1465–1472, 2003.
- [18] G. Shani, D. Heckerman, and R. I. Brafman, "An mdp-based recommender system," *Journal of Machine Learning Research*, vol. 6, no. Sep, pp. 1265–1295, 2005.
- [19] T. Hofmann, "Collaborative filtering via gaussian probabilistic latent semantic analysis," in *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pp. 259–266, ACM, 2003.
- [20] T. Hofmann, "Latent semantic models for collaborative filtering," *ACM Transactions on Information Systems (TOIS)*, vol. 22, no. 1, pp. 89–115, 2004.
- [21] B. M. Marlin, "Modeling user rating profiles for collaborative filtering," in *Advances in neural information processing systems*, pp. 627–634, 2004.
- [22] Y. Koren, "Factorization meets the neighborhood: a multifaceted collaborative filtering model," in *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 426–434, ACM, 2008.
- [23] P. Cremonesi, Y. Koren, and R. Turrin, "Performance of recommender algorithms on top-n recommendation tasks," in *Proceedings of the fourth ACM conference on Recommender systems*, pp. 39–46, ACM, 2010.
- [24] X. He, T. Chen, M.-Y. Kan, and X. Chen, "Trirank: Review-aware explainable recommendation by modeling aspects," in *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pp. 1661–1670, ACM, 2015.
- [25] B. Marlin, R. S. Zemel, S. Roweis, and M. Slaney, "Collaborative filtering and the missing at random assumption," *arXiv preprint arXiv:1206.5267*, 2012.
- [26] B. Hidasi, A. Karatzoglou, L. Baltrunas, and D. Tikk, "Session-based recommendations with recurrent neural networks," *arXiv preprint arXiv:1511.06939*, 2015.
- [27] I. Bayer, X. He, B. Kanagal, and S. Rendle, "A generic coordinate descent framework for learning from implicit feedback," in *Proceedings of the 26th International Conference on World Wide Web*, pp. 1341–1350, International World Wide Web Conferences Steering Committee, 2017.
- [28] J. Chen, H. Zhang, X. He, L. Nie, W. Liu, and T.-S. Chua, "Attentive collaborative filtering: Multimedia recommendation with item-and component-level attention," in *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 335–344, ACM, 2017.
- [29] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, "Bpr: Bayesian personalized ranking from implicit feedback," in *Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence*, pp. 452–461, AUAI Press, 2009.

- [30] X. He, H. Zhang, M.-Y. Kan, and T.-S. Chua, “Fast matrix factorization for online recommendation with implicit feedback,” in *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pp. 549–558, ACM, 2016.
- [31] A. Anderson, R. Kumar, A. Tomkins, and S. Vassilvitskii, “The dynamics of repeat consumption,” in *Proceedings of the 23rd international conference on World wide web*, pp. 419–430, ACM, 2014.
- [32] Y. Koren, “Collaborative filtering with temporal dynamics,” *Communications of the ACM*, vol. 53, no. 4, pp. 89–97, 2010.
- [33] T. Miranda, M. Claypool, A. Gokhale, T. Mir, P. Murnikov, D. Netes, and M. Sartin, “Combining content-based and collaborative filters in an online newspaper,” in *In Proceedings of ACM SIGIR Workshop on Recommender Systems*, Citeseer, 1999.
- [34] M. J. Pazzani, “A framework for collaborative, content-based and demographic filtering,” *Artificial intelligence review*, vol. 13, no. 5-6, pp. 393–408, 1999.
- [35] M. Balabanović and Y. Shoham, “Fab: content-based, collaborative recommendation,” *Communications of the ACM*, vol. 40, no. 3, pp. 66–72, 1997.
- [36] I. Soboroff and C. Nicholas, “Combining content and collaboration in text filtering,” in *Proceedings of the IJCAI*, vol. 99, pp. 86–91, sn, 1999.
- [37] C. Basu, H. Hirsh, W. Cohen, *et al.*, “Recommendation as classification: Using social and content-based information in recommendation,” in *Aaai/iaai*, pp. 714–720, 1998.
- [38] A. Popescul, D. M. Pennock, and S. Lawrence, “Probabilistic models for unified collaborative and content-based recommendation in sparse-data environments,” in *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*, pp. 437–444, Morgan Kaufmann Publishers Inc., 2001.
- [39] A. I. Schein, A. Popescul, L. H. Ungar, and D. M. Pennock, “Methods and metrics for cold-start recommendations,” in *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 253–260, ACM, 2002.
- [40] J. Madhusudanan, A. Selvakumar, and R. Sudha, “Frame work for context aware applications,” in *Computing Communication and Networking Technologies (ICCCNT), 2010 International Conference on*, pp. 1–4, IEEE, 2010.
- [41] K. Meehan, T. Lunney, K. Curran, and A. McCaughey, “Context-aware intelligent recommendation system for tourism,” in *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2013 IEEE International Conference on*, pp. 328–331, IEEE, 2013.
- [42] M. Benerecetti, P. Bouquet, and M. Bonifacio, “Distributed context-aware systems,” *Human-Computer Interaction*, vol. 16, no. 2, pp. 213–228, 2001.
- [43] T. Mahmood, F. Ricci, and A. Venturini, “Improving recommendation effectiveness: Adapting a dialogue strategy in online travel planning,” *Information Technology & Tourism*, vol. 11, no. 4, pp. 285–302, 2009.
- [44] A. Karatzoglou, X. Amatriain, L. Baltrunas, and N. Oliver, “Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering,” in *Proceedings of the fourth ACM conference on Recommender systems*, pp. 79–86, ACM, 2010.
- [45] Y. Sun, N. J. Yuan, X. Xie, K. McDonald, and R. Zhang, “Collaborative nowcasting for contextual recommendation,” in *Proceedings of the 25th International Conference on World Wide Web*, pp. 1407–1418, International World Wide Web Conferences Steering Committee, 2016.
- [46] C. Lei-hui, K. Jun, C. Hui, Z. Wei, Z. Jian-bing, and G. Ming, “Techniques for cross-domain recommendation: A survey,” 2017.
- [47] M. Jiang, P. Cui, F. Wang, Q. Yang, W. Zhu, and S. Yang, “Social recommendation across multiple relational domains,” in *Proceedings of the 21st ACM international conference on Information and knowledge management*, pp. 1422–1431, ACM, 2012.

- [48] S. Gao, H. Luo, D. Chen, S. Li, P. Gallinari, and J. Guo, “Cross-domain recommendation via cluster-level latent factor model,” in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 161–176, Springer, 2013.
- [49] G. D. Montanez, R. W. White, and X. Huang, “Cross-device search,” in *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pp. 1669–1678, ACM, 2014.
- [50] M. Jiang, P. Cui, X. Chen, F. Wang, W. Zhu, and S. Yang, “Social recommendation with cross-domain transferable knowledge,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 11, pp. 3084–3097, 2015.
- [51] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [52] A. Van den Oord, S. Dieleman, and B. Schrauwen, “Deep content-based music recommendation,” in *Advances in neural information processing systems*, pp. 2643–2651, 2013.
- [53] H. Wang, N. Wang, and D.-Y. Yeung, “Collaborative deep learning for recommender systems,” in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1235–1244, ACM, 2015.
- [54] S. Li, J. Kawale, and Y. Fu, “Deep collaborative filtering via marginalized denoising auto-encoder,” in *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pp. 811–820, ACM, 2015.
- [55] X. Geng, H. Zhang, J. Bian, and T.-S. Chua, “Learning image and user features for recommendation in social networks,” in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4274–4282, 2015.
- [56] X. Dong, L. Yu, Z. Wu, Y. Sun, L. Yuan, and F. Zhang, “A hybrid collaborative filtering model with deep structure for recommender systems,” in *AAAI*, pp. 1309–1315, 2017.
- [57] H. Guo, R. Tang, Y. Ye, Z. Li, and X. He, “Deepfm: A factorization-machine based neural network for ctr prediction,” *arXiv preprint arXiv:1703.04247*, 2017.
- [58] H.-T. Cheng, L. Koc, J. Harmsen, T. Shaked, T. Chandra, H. Aradhye, G. Anderson, G. Corrado, W. Chai, M. Ispir, *et al.*, “Wide & deep learning for recommender systems,” in *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems*, pp. 7–10, ACM, 2016.
- [59] P. Covington, J. Adams, and E. Sargin, “Deep neural networks for youtube recommendations,” in *Proceedings of the 10th ACM Conference on Recommender Systems*, pp. 191–198, ACM, 2016.
- [60] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T.-S. Chua, “Neural collaborative filtering,” in *Proceedings of the 26th International Conference on World Wide Web*, pp. 173–182, International World Wide Web Conferences Steering Committee, 2017.
- [61] P.-S. Huang, X. He, J. Gao, L. Deng, A. Acero, and L. Heck, “Learning deep structured semantic models for web search using clickthrough data,” in *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, pp. 2333–2338, ACM, 2013.
- [62] H.-J. Xue, X.-Y. Dai, J. Zhang, S. Huang, and J. Chen, “Deep matrix factorization models for recommender systems,” *static. ijcai. org*, 2017.
- [63] S. Wold, K. Esbensen, and P. Geladi, “Principal component analysis,” *Chemometrics and intelligent laboratory systems*, vol. 2, no. 1-3, pp. 37–52, 1987.
- [64] A. J. Izenman, “Linear discriminant analysis,” in *Modern multivariate statistical techniques*, pp. 237–280, Springer, 2013.
- [65] F. Wickelmaier, “An introduction to mds,” *Sound Quality Research Unit, Aalborg University, Denmark*, vol. 46, 2003.
- [66] J. B. Tenenbaum, V. De Silva, and J. C. Langford, “A global geometric framework for nonlinear dimensionality reduction,” *science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

- [67] S. T. Roweis and L. K. Saul, “Nonlinear dimensionality reduction by locally linear embedding,” *science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [68] M. Belkin and P. Niyogi, “Laplacian eigenmaps and spectral techniques for embedding and clustering,” in *Advances in neural information processing systems*, pp. 585–591, 2002.
- [69] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, “Distributed representations of words and phrases and their compositionality,” in *Advances in neural information processing systems*, pp. 3111–3119, 2013.
- [70] X. Wang, P. Cui, J. Wang, J. Pei, W. Zhu, and S. Yang, “Community preserving network embedding,” in *AAAI*, pp. 203–209, 2017.
- [71] X. Zhao, A. Chang, A. D. Sarma, H. Zheng, and B. Y. Zhao, “On the embeddability of random walk distances,” *Proceedings of the VLDB Endowment*, vol. 6, no. 14, pp. 1690–1701, 2013.
- [72] T. Man, H. Shen, S. Liu, X. Jin, and X. Cheng, “Predict anchor links across social networks via an embedding approach,” in *IJCAI*, pp. 1823–1829, 2016.
- [73] S. Cao, W. Lu, and Q. Xu, “Grarep: Learning graph representations with global structural information,” in *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pp. 891–900, ACM, 2015.
- [74] S. Cao, W. Lu, and Q. Xu, “Deep neural networks for learning graph representations,” in *AAAI*, pp. 1145–1152, 2016.
- [75] L. Liu, W. K. Cheung, X. Li, and L. Liao, “Aligning users across social networks using network embedding,” in *IJCAI*, pp. 1774–1780, 2016.
- [76] C. Zhou, Y. Liu, X. Liu, Z. Liu, and J. Gao, “Scalable graph embedding for asymmetric proximity,” in *AAAI*, pp. 2942–2948, 2017.
- [77] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei, “Line: Large-scale information network embedding,” in *Proceedings of the 24th International Conference on World Wide Web*, pp. 1067–1077, International World Wide Web Conferences Steering Committee, 2015.
- [78] A. Grover and J. Leskovec, “node2vec: Scalable feature learning for networks,” in *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 855–864, ACM, 2016.
- [79] S. Wang, J. Tang, C. Aggarwal, Y. Chang, and H. Liu, “Signed network embedding in social media,” in *Proceedings of the 2017 SIAM International Conference on Data Mining*, pp. 327–335, SIAM, 2017.
- [80] S. Yuan, X. Wu, and Y. Xiang, “Sne: Signed network embedding,” in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp. 183–195, Springer, 2017.
- [81] S. Chang, W. Han, J. Tang, G.-J. Qi, C. C. Aggarwal, and T. S. Huang, “Heterogeneous network embedding via deep architectures,” in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 119–128, ACM, 2015.
- [82] Q. Wang, Z. Mao, B. Wang, and L. Guo, “Knowledge graph embedding: A survey of approaches and applications,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 12, pp. 2724–2743, 2017.
- [83] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko, “Translating embeddings for modeling multi-relational data,” in *Advances in neural information processing systems*, pp. 2787–2795, 2013.
- [84] Z. Wang, J. Zhang, J. Feng, and Z. Chen, “Knowledge graph embedding by translating on hyperplanes,” in *AAAI*, pp. 1112–1119, 2014.
- [85] Y. Lin, Z. Liu, M. Sun, Y. Liu, and X. Zhu, “Learning entity and relation embeddings for knowledge graph completion,” in *AAAI*, pp. 2181–2187, 2015.
- [86] J. Li, J. Zhu, and B. Zhang, “Discriminative deep random walk for network classification,” in *ACL (1)*, 2016.

- [87] M. Niepert, M. Ahmed, and K. Kutzkov, “Learning convolutional neural networks for graphs,” in *International Conference on Machine Learning*, pp. 2014–2023, 2016.
- [88] C. Yang, Z. Liu, D. Zhao, M. Sun, and E. Y. Chang, “Network representation learning with rich text information,” in *IJCAI*, pp. 2111–2117, 2015.
- [89] T. M. Le and H. W. Lauw, “Probabilistic latent document network embedding,” in *Data Mining (ICDM), 2014 IEEE International Conference on*, pp. 270–279, IEEE, 2014.
- [90] P. Yanardag and S. Vishwanathan, “Deep graph kernels,” in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1365–1374, ACM, 2015.
- [91] N. Shervashidze, P. Schweitzer, E. J. v. Leeuwen, K. Mehlhorn, and K. M. Borgwardt, “Weisfeiler-lehman graph kernels,” *Journal of Machine Learning Research*, vol. 12, no. Sep, pp. 2539–2561, 2011.
- [92] B. Perozzi, R. Al-Rfou, and S. Skiena, “Deepwalk: Online learning of social representations,” in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 701–710, ACM, 2014.
- [93] F. Tian, B. Gao, Q. Cui, E. Chen, and T.-Y. Liu, “Learning deep representations for graph clustering,” in *AAAI*, pp. 1293–1299, 2014.
- [94] D. Wang, P. Cui, and W. Zhu, “Structural deep network embedding,” in *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1225–1234, ACM, 2016.
- [95] F. Zhang, N. J. Yuan, D. Lian, X. Xie, and W.-Y. Ma, “Collaborative knowledge base embedding for recommender systems,” in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 353–362, ACM, 2016.
- [96] H. Wang, F. Zhang, M. Hou, X. Xie, M. Guo, and Q. Liu, “Shine: Signed heterogeneous information network embedding for sentiment link prediction,” *arXiv preprint arXiv:1712.00732*, 2017.
- [97] H. Wang, J. Wang, J. Wang, M. Zhao, W. Zhang, F. Zhang, X. Xie, and M. Guo, “Graphgan: Graph representation learning with generative adversarial nets,” *arXiv preprint arXiv:1711.08267*, 2017.