

《智能信息处理》课程考试

推荐系统应用中的多模态知识图谱构建

覃文强

考核	到课[10]	作业[20]	考试[70]	课程成绩[100]
得分				

2021 年 12 月 10 日

推荐系统应用中的多模态知识图谱构建

覃文强

(大连海事大学 信息科学技术学院, 辽宁省大连市 中国 116026)

摘要 知识图谱将知识库以一种图谱的形式展现出来, 使知识具有可解释性、可推理性, 从而使机器具备认知能力, 是人工智能的重要基石。近年来, 知识图谱 (Knowledge Graph) 与推荐系统结合的工作成为了研究的热门方向。现如今大部分基于知识图谱的推荐算法都从学习知识图谱中的语义信息或探索知识图谱中各种路径的连接模式两个方面来利用知识图谱增强推荐算法的效果, 而忽略了本身就存在的并且可以利用的多模态信息, 合理的利用这些信息可能会给推荐算法的性能带来提升。多模态知识图谱是将多模态信息引入到知识图谱的一种技术, 因此可以作为推荐系统的知识源。在很多推荐场景中, 比如电影推荐场景中的导演, 演员信息; 音乐推荐场景中的歌手, 乐队信息等, 可以利用这些常识信息构建知识图谱。在此基础上, 图像视觉知识, 文本知识等多模态信息可以将知识图谱丰富为多模态知识图谱, 这样可以很自然的让基于知识图谱的推荐系统利用到多模态信息从而提升推荐效果。基于上述背景, 本文会对推荐系统应用中的多模态知识图谱的构建进行讨论。

关键词 知识图谱; 多模态; 推荐算法

中图法分类号 TP311.20 DOI 号 10.3969/j.issn.1001-3695.2014.01.030

Construction of multimodal knowledge graph in the application of recommendation system

Qin Wenqiang

(School of Information Science and Technology, Dalian maritime university, Liaoning Dalian)

Abstract The knowledge graph shows the knowledge base in the form of a graph, which makes the knowledge explainable and inferential, so that the machine has cognitive ability, which is an important cornerstone of artificial intelligence. In recent years, the combination of knowledge graph and recommendation system has become a hot research direction. Nowadays, most recommendation algorithms based on knowledge graph use knowledge graph to enhance the effect of recommendation algorithm from two aspects: learning semantic information in knowledge graph or exploring the connection mode of various paths in knowledge graph. While ignoring the existing and available multimodal information, reasonable use of this information may improve the performance of the recommendation algorithm. Multimodal knowledge graph is a technology that introduces multimodal information into knowledge graph, so it can be used as the knowledge source of recommendation system. In many recommended scenes, such as the director and actor information in the movie recommendation scene, the singer and band information in the music recommendation scene, we can use these common sense information to build a knowledge graph. On this basis, the multimodal information such as image visual knowledge and text knowledge can enrich the knowledge graph into the multimodal knowledge graph, which can naturally make the recommendation system based on the knowledge graph use the multimodal information to improve the recommendation effect. Based on the above background, this paper will discuss the construction of multimodal knowledge graph in the application of recommendation system.

Keywords Knowledge Graph; Multimodal; Recommendation Algorithm

1 引言

随着信息技术的快速发展,推荐系统逐渐成为了很多互联网产品的核心技术。推荐算法解决的核心问题就是代替用户来评估用户本身没有消费过的商品或服务,并将用户可能感兴趣的商品或服务推荐给用户。在大型信息或者电商网站,用户点击或者消费过的物品总是占总物料池的冰山一角,因此推荐算法会遇见数据稀疏(Data Sparsity)问题。为了缓解数据稀疏的问题,研究者们通过引入物品的辅助信息(Side Information)来增强物品或者用户的表征能力。因为除了用户和物品的历史交互信息可以利用以外,用户和物品一般携带一些辅助信息,比如物品的各种属性(类别,标签),用户的基本信息(性别,年龄),合理的利用辅助信息可以有效的缓解此类问题^[1]近年来,知识图谱(Knowledge Graph)与推荐系统结合的工作成为了研究的热门方向。

在推荐系统中,物品侧是很容易利用已有信息构建知识图谱的,比如电影推荐场景,电影会有导演,演员等信息;书籍推荐场景中,书籍会有出版社,作者等等信息,可以很自然的使用这些知识组成的知识图谱来辅助推荐算法。知识图谱的引入在提供了知识的同时还可以增加不同物品之间的联系,让原本本身没有关联的两个物品通过知识图谱的中间实体节点有了新的连接,换句话说就是强化了商品之间的联系,从而可以更准确地捕获用户偏好和物品之间的隐藏联系,进而提升推荐算法的整体性能。

随着网络速度的提升,多元化的信息逐渐走进人们的生活,例如音视频等媒体信息。随之而来的是越来越多的信息类型可以使用在推荐算法中,我们一般称多种类型的数据集合为多模态数据。传统的基于知识图谱的推荐算法忽略了多模态信息,而多模态信息往往参与用户的决策过程,比如本地生活平台大众点评,用户习惯在去一家店消费之前先查看一下店铺的评论信息(文本模态),以及对应的用户发布的店铺相关的图片信息(图像模态)。同样的,在电影推荐场景下,用户可能习惯于去看一部电影前先去看一下该电影的影评(文本模态),该电影的宣传片或宣传海报(图像模态),在查看宣传片的时候甚至有些用户是被该电影的配乐吸引而去观看该电影的(声音模态)。现如今大部分

基于知识图谱的推荐算法都从学习知识图谱中的语义信息^[2-4]或探索知识图谱中各种路径的连接模式两个方面来^[5,6]利用知识图谱增强推荐算法的效果,而忽略了这些本身就存在的并且可以利用的多模态信息,合理的利用这些信息可能会给推荐算法的性能带来提升。基于上述背景,本文会对多模态知识图谱的应用进行讨论。

2 多模态知识图谱的概念

我们目前了解的部分机器学习算法或者深度学习算法,无论是监督学习还是非监督学习,都是从现有的数据中学习一些模式。这些方式或是从关联规则的角度进行学习,或是从统计概率的角度进行学习,这些学习方式也是符合人类的直觉的。但是我们可以思考一下人类学习的过程,人类在学习一种新的知识时,通常不仅仅局限于当前学习的内容,而且会思考或调用一些在之前的学习过程中学到的知识,这些知识可以理解为是先验知识或者是常识。如果让计算机在学习过程中也可以引入一些这样的先验或者常识,能否提升计算机或者机器学习算法的学习能力呢?在这样的技术背景下,知识图谱^[7]被研究者们所提出。

知识图谱是将知识组织成图的形式,方便计算机的查询和理解。知识图谱就是知识在计算机或者机器学习算法中知识的存在形式。知识图谱在很多领域都有应用,比如信息检索,我们在百度上简单的搜索“姚明的妻子是谁?”,百度就会直接返回给我们一个精确的结果是“叶莉”,同时附上相应的人物信息,比如职业,年龄等人物的基本信息,而不是像其他的百度搜索场景那样返回给我们一个排序后的链接地址。这是因为百度搜索系统在内部维护了一个名人关系的知识图谱,其中包含人物的基本信息和与其相关的人或者物品,比如亲属,工作单位,所在地等信息。在信息检索的过程中,通过图数据库技术可以很快的检索出对应的知识。通过这种方式来提升人们使用搜索引擎的体验。除了信息检索领域,还有很多方向使用到了知识图谱,例如智能问答,企业风控,推荐系统等方向。

知识图谱是知识在计算机中的表达形式,通常以结构化的形式存在,从图谱的角度理解,结构化的知识代表逻辑的,可以推理的知识。但是知识不仅仅可以从推理的角度理解,对于人类来说,有一些知识是可以从感官上直接获取的知识,这些知识

的特点是直观的,快速的,无意识的,非语言特点的,比如图片数据(视觉知识),人在观察图片的时候就会摄取到类似的知识。比如在观察成都太古里的图片的时候,用户会在心里给出图片直接的理解,可以是建筑特色,城市风情等,而且此类理解很难被结构化的知识所表达。同样的,声音知识也是如此。知识图谱中一些实体本身就含有视觉或其他多模态的信息,不同模态的信息对于知识的补充各有侧重,有助于从不同的方面来增强知识的表达。同时,现有技术缺乏有效手段将视觉知识提取成结构化的知识加入到知识图谱的三元组中来,得益于深度学习的发展,研究者们探索了对于多种模态数据的表征方法,例如计算机视觉(CV)和自然语言处理(NLP)就分别对图像与文本进行表征,并已经有了非常迅速的发展。我们可以利用这些技术直接对视觉知识和文本知识进行表征,直接在知识图谱中利用视觉和文本知识。将多模态数据引入到知识图谱而构造的新的图谱就叫做多模态知识图谱^[8],一个简单的电影多模态知识图谱的实例如图 1-2 所示。

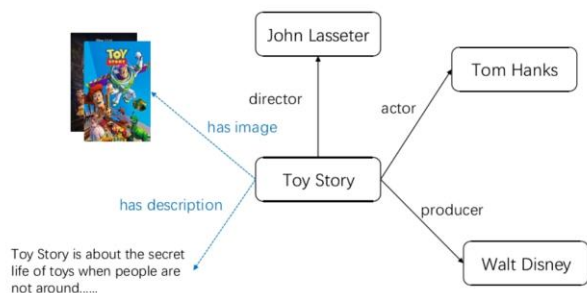


图 2-1 一个多模态知识图谱的例子

3 多模态知识图谱构建

3.1 多模态知识图谱数据集构建

因为目前多模态知识图谱数据比较少,而且基本都存在于通用领域。但是在推荐系统场景下,多模态信息是可以获得的,因此本节会先介绍一下多模态知识图谱数据集是如何构建的。为了体现泛化性,本文选择了两个差异性比较大的推荐场景进行实验,两个场景分别代表了两个推荐领域:电影推荐和餐厅推荐。接下来会分别介绍两个场景的数据集是如何进行构建的。

3.1.1 影视多模态知识图谱构建

MovieLens 经常出现在推荐系统的相关论文中,以用于评估推荐算法的优劣。其中,MovieLens

数据集中的每一条数据是用户对于已经看过的影视片的一个评级,其中评级的分数范围为 1-5,5 分为评价最高,1 分为评价最低,其余的分数以此类推。首先本文要构建推荐系统任务的数据集,因为 MovieLens 数据集给定的是用户对于看过的影视片的评级,但是对于一般的推荐场景来说,建模的问题是点击率预估,也就是二分类问题,因此本文需要将用户对于电影的评级转换为隐式反馈数据,也就是用户是否观看。具体地,因为用户打分过的影视一定是用户看过的,所以本文将用户已经评分的影视片标记为 1(正例),对于没有评分的影视片进行随机采样,标记为 0(负例)。

MovieLens 影视知识图谱的结构化的知识图谱数据来源于工作 KGNN-LS^[10],KGNN-LS 利用微软的 Satori 知识图谱作为数据源来构建 MovieLens 的影视知识图谱,可以理解为 Satori 知识图谱是本文使用的影视知识图谱的父图谱。具体地,首先从 Satori 知识图谱中选择置信度高于 0.9 的集合,记为集合 S。KGNN-LS 使用集合 S,通过将 MovieLens 数据集中的电影的名称与集合 S 中三元组的尾实体进行匹配来收集所有有效电影的 Satori ID。拥有一组物品 ID 后,KGNN-LS 再将这些电影 ID 与集合 S 中所有三元组的头实体匹配,并选择可以匹配的三元组作为最终的知识图谱。

接下来要构建 MovieLens 影视知识图谱的多模态信息,首先为了构建 Movie-Lens 影视知识图谱的图像实体,本文从 Youtubex 网站上抓取实体对应的宣传片,具体的宣传片的抓取使用 Github 的开源工具 PyTubey。有了宣传片后,然后使用 FFmpegz(视频编解码处理软件)来抽取每一个宣传片的关键帧,使用这些关键帧作为视觉知识,最后使用预训练的 ResNet50 模型来抽取关键帧的视觉特征。接下来构建 MovieLens 影视知识图谱的文本知识。本文从 TMDB 中抓取每一个视频对应的影视描述,使用影视描述作为该多模态知识图谱的文本知识,然后使用 SIF 句向量模型制作句向量来表达文本特征。

3.1.2 美食多模态知识图谱构建

Dianping[1],大众点评是来自中国的生活信息服务网站,用户可以在其中搜索和获取餐厅,酒店,机票,景点等信息,本文选择了美食场景下的餐厅推荐作为推荐场景,Dianping 数据集由 Meituan-Dianping Group 提供。首先要构建推荐系统任务的

数据集,在 Dianping 数据中,正例包括购买和加入收藏夹;负例是从每一个用户随机采样而来(采样该用户历史没有交互过的数据)。Dianping 知识图谱是从美团大脑收集的,这是由 Meituan-Dianping Group 为餐饮和娱乐而建立的知识图谱。Dianping 知识图谱的实体类型包括 POI(即餐馆),餐厅的一级分类和二级分类(比如川菜,西餐,日本餐等),餐厅所在地区(通常为各种商圈)和标签(类似于推荐系统中的标签系统,例如口味好,环境好,服务好等标签),星级等。为了构造 Dianping 知识图谱的图像实体,本文选择了 POI 的头部推荐菜品的图像(按照用户推荐的顺序排序,排在前面的为头部)。

表 3-1 数据集的统计信息

dataset	MovieLens	Dianping
# of users	41849	40388
# of items	4828	29969
# of interactions	1813381	624499
# of entities	65801	93798
# of relations	19	6
# of triplets	145406	635656

与 MovieLens 数据集类似,本文使用预先训练好的 ResNet50^[9]模型从推荐菜的图像中提取视觉特征。为了构造 Dianping 知识图谱的文本实体,本文使用了每一个 POI 中的 UGC(用户评论)作为文本知识,这些 UGC 经过数据清洗后使用。本文的美食多模态知识图谱是依托 Dianping 数据集中物品的知识所构建的,所以在本文中基于美食多模态知识图谱的推荐系统数据集简称为 Dianping。其中,两个数据集的一些统计信息展示在表 3-1,主要从本文模型中的两个不同的任务的角度进行统计。推荐系统任务方面:主要统计了用户、物品数量,用户物品交互数量。知识图谱方面,统计了两个数据集中知识图谱的各项统计指标:三元组个数,实体和关系分别的个数。

3.2 协同多模态知识图谱构建

3.2.1 知识图谱

知识图谱(Knowledge Graph)^[6]是由实体(节点)和关系(含有类型的边)组成的异质信息网络,其具体定义为 $G_{kg} = (V_{kg}, E_{kg})$, 其中 V_{kg} 表示节点

的集合, E_{kg} 表示边的集合。一般来说节点是实体(Entity),边通过连接两个实体来构成一个具有主体,属性,对象的三元组事实。每一个边都属于一个关系类型 $r \in R_{kg}$, R_{kg} 是关系类型的集合。知识图谱中每一个知识的最小单位是实体的三元组,以(头部实体,关系,尾部实体)的形态存在,一般表示为 (h, r, t) , 其中 $h, t \in E_{kg}$, $r \in R_{kg}$ 。例如(Chengdu, capital of Sichuan)表示成都是四川的省会。许多实体通过关系相连接,就构成了图结构,如图 3-1 所示,因为某些中间节点可以将不同领域的知识图谱相连接从而形成了更大的知识图谱,这个图是一个简单的知识图谱例子。知名的知识图谱包括 FreeBase^[11], WikiData^[12], DBPedia^[13], 其中有些规模比较大的知识图谱可以包含百万级别的实体和亿级别的事实,因此可以说大型的知识图谱包含海量的知识也不为过。

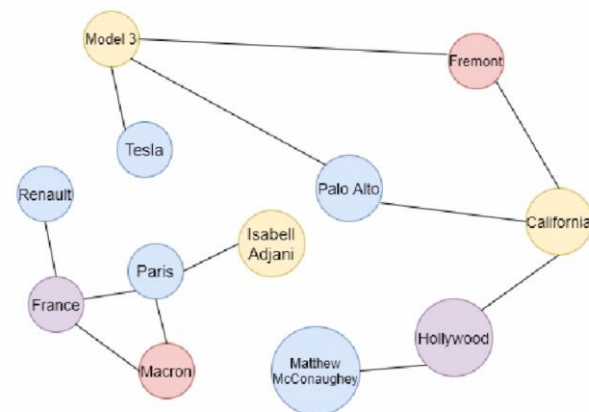


图 3-1 知识图谱的图结构

多模态知识图谱也是广义知识图谱的一种,从本质上看没有太大的区别。多模态知识图谱是在原有知识图谱的基础上引入了其他类型(模态)的信息,比如文本模态或者图像模态。本文所使用的多模态知识图谱指的是基于实体的方式的多模态知识图谱,因为此类方式对于多模态的数据源比较友好,数据可以很方便的获取。一个简单的基于实体的多模态知识图谱的示意图可以见第一章中的图 2-1,从图中可以看出,基于实体的多模态知识图谱是通过引入新的关系,让多模态信息作为节点存在知识图谱中,方便下游任务的使用。在推荐系统领域经常会有图(Graph / Network)结构的数据,比如常用的标签系统,分类系统等。工程师或研究者们会使用一些算法模型来利用这些图结构的数据到推荐系统模型中来,比如使用 Deepwalk^[14],

Line^[15]等网络嵌入算法将每一个物品或者用户的标签转化为嵌入向量再入模。图结构的数据除了使用图嵌入算法转换为特征使用之外,还可以直接使用,因为用户和其历史交互过的物品先天可以构成图结构,这种图叫做用户-物品交互二部图(User-Item Bipartite),近年来,也有一些直接利用用户-物品交互二部图进行推荐的工作出现,比如基于图的矩阵分解 GCMC^[16],基于图神经网络的协同过滤算法 NGCF^[17]等,这些算法的特点就是可以让算法不局限于学习现有的数据的标签的模式,而是加入了数据连接等信息进行推荐,也取得了不错的效果。

3.2.2 用户-物品二部图

用户-物品交互二部图其具体定义为 $G_{bi} = (V_{bi}, E_{bi})$, 其中 V_{bi} 表示节点的集合, E_{bi} 表示边的集合。假定物品集合为 $i \in I$ 用户集合为 $u \in U$, 那么在 G_{bi} 中, G_{bi} 定义每个用户和每个物品都是节点, 即 $i \in V_{bi}, u \in V_{bi}$ 。 G_{bi} 中的边 y_{bi} 为用户和物品的历史交互, 即 $y_{bi} \in E_{bi}$, 当用户 u 和物品 i 有过历史交互时, $y_{bi}=1$, 否则的话 $y_{bi}=0$, 一般来说 G_{bi} 只保留 $y_{bi}=1$ 的边, 也有一些工作将边的权重设计的更为复杂, 比如使用用户历史评分作为边的权重。用户-物品交互二部图的一个简单实例如图 3-2 所示, 图中对于用户和物品节点, 使用了不同颜色的圆点来表示。

3.2.3 协同多模态知识图谱构建

协同多模态知识图谱是混合用户物品二部图和多模态知识图谱而形成的一个新的异质信息网络, 也属于广义知识图谱的一种, 接下来会给出其详细定义。协同多模态知识图谱首先将之前的用户

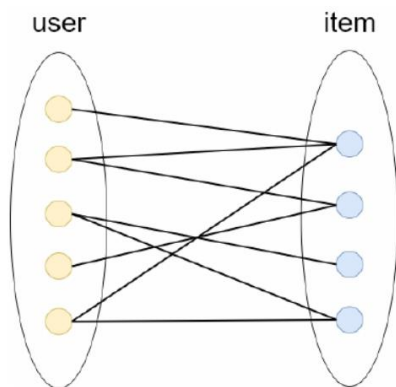


图 3.2 用户-物品交互二部图

物品交互二部图中的物品和用户节点都定义为实体, 具体为 $\{(e_u, y_{ui}, e_i) | e_u \in U, e_i \in I\}$, 其中

e_u 表示用户实体, y_{ui} 表示用户 u 和物品 i 之间的连接, e_i 表示物品实体, U and I 分别表示用户和物品集。当用户实体 e_u 和物品实体 e_i 有过历史交互时, $y_{ui}=1$, 否则的话 $y_{ui}=0$ 。然后协同多模态知识图谱会将多模态知识图谱合并到用户-物品二部图中。具体来说, 本文将 $y_{ui}=1$ 的边保留, 将这些用户行为表示成知识图谱中的三元组的形式 $(e_u, \text{Interact}, e_i)$, 定义用户-物品的边为关系 **Interact**。同时, 物品会含有一些知识, 经由这些知识构建知识图谱。因为二部图和知识图谱都有物品结点, 所以基于物品-实体对齐的集合, 用户-物品二部图可以与知识图谱合并, 形成一个统一的图。如图 3-3 所示, e_{i1} , e_{i2} 和 e_{i3} 三个实体都既出现再用户-物品二部图中, 又出现再知识图谱中, 因此协同知识图谱可以通过它们实现知识图谱和用户-物品交互二部图的对齐, 图 3-3 为协同多模态知识图谱的一个样例, 图中使用不同颜色的节点表示不同的实体类型, 使用不同的线条表示不同的关系类型, 具体可以见图例。遵循以上构建步骤, 即可建立模型所使用的协同多模态知识图谱。

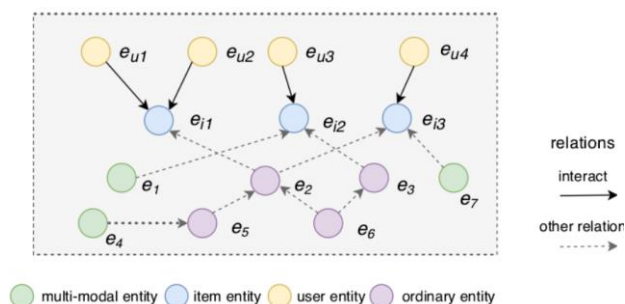


图 2.1 协同多模态知识图谱的一个样例

结束语

首先介绍了本文所使用的含有多模态知识图谱数据集的构建过程, 分别从推荐系统数据集构造, 知识图谱中结构化知识获取以及多模态信息获取三个方面介绍。其中两个数据集的构建过程其实大同小异, 也适合扩展到其他领域。然后介绍了本文所使用的用户-物品二部图的构造过程和本文提出的模型输入, 即协同多模态知识图谱的构造过程。

参 考 文 献

- [1] Q. Guo, F. Zhuang, C. Qin, et al. A survey on knowledge graph-based recommender systems[J]. IEEE Transactions on Knowledge and Data Engineering, 2020
- [2] H. Wang, F. Zhang, M. Hou, et al. SHINE: signed heterogeneous information network embedding for sentiment link prediction[C]. Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, 2018, 592-600
- [3] H. Wang, F. Zhang, X. Xie, et al. DKN: deep knowledge-aware network for news recommendation[C]. Proceedings of the 2018 World Wide Web Conference on World Wide Web, 2018, 1835-1844
- [4] Q. Wang, Z. Mao, B. Wang, et al. Knowledge graph embedding: A survey of approaches and applications[J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(12): 2724-2743
- [5] X. Yu, X. Ren, Y. Sun, et al. Personalized entity recommendation: a heterogeneous information network approach[C]. Seventh ACM International Conference on Web Search and Data Mining, 2014, 283-292
- [6] H. Zhao, Q. Yao, J. Li, et al. Meta-graph based recommendation fusion over heterogeneous information networks[C]. Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2017, 635-644
- [7] S. Ji, S. Pan, E. Cambria, et al. A survey on knowledge graphs: Representation, acquisition and applications[J]. arXiv preprint arXiv:2002.00388, 2020
- [8] Y. Liu, H. Li, A. Garcia-Duran, et al. Mmkg: multi-modal knowledge graphs[C]. European Semantic Web Conference, 2019, 459-474
- [9] K. He, X. Zhang, S. Ren, et al. Deep residual learning for image recognition[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016, 770-778
- [10] H. Wang, F. Zhang, M. Zhang, et al. Knowledge-aware graph neural networks with label smoothness regularization for recommender systems[C]. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019, 968-977
- [11] K. Bollacker, C. Evans, P. Paritosh, et al. Freebase: a collaboratively created graph database for structuring human knowledge[C]. Proceedings of the 2008 ACM SIGMOD international conference on Management of data, 2008, 1247-1250
- [12] D. Vrandečić, M. Krötzsch. Wikidata: a free collaborative knowledgebase[J]. Communications of the ACM, 2014, 57(10): 78-85
- [13] S. Auer, C. Bizer, G. Kobilarov, et al. Dbpedia: A nucleus for a web of open data[M]. Springer, 2007, 722-735
- [14] B. Perozzi, R. Al-Rfou, S. Skiena. Deepwalk: online learning of social representations[C]. The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2014, 701-710
- [15] J. Tang, M. Qu, M. Wang, et al. LINE: large-scale information network embedding[C]. Proceedings of the 24th International Conference on World Wide Web, 2015, 1067-1077