

AI 前沿讲座刘康老师作业

基于知识图谱的推荐系统

学生： 槐 泽 鹏

导师： X X X

学号:xxxxxxxxxxxxxxxxxxxx

联系方式: xxxxxxxxxxxxxxx

2021 年 11 月 3 日

目 录

第 1 节 背景	1
1.1 推荐	1
1.2 知识图谱与推荐	1
第 2 节 基于知识图谱的推荐算法	2
2.1 基于 KGE 的算法	2
2.2 基于知识图谱路径的方法	6
2.3 基于 GNN 的算法	8
第 3 节 小结	11
第 4 节 参考文献	12

第 1 节 背景

本报告旨在完成 2020 年人工智能前沿讲座选做大作业中刘康老师布置的作业：“以大报告的形式，就知识图谱某一项前沿技术写出自己的认知和体会”。

本人博士研究方向为“图数据挖掘和基于图的推荐系统”，知识图谱是天然的图结构化数据，“基于知识图谱的推荐系统”是近些年“图推荐”的热点。因此本报告拟对“知识图谱推荐”做一个综述，同时也是近期调研工作的一个总结。

1.1 推荐

推荐系统是指为用户预测其感兴趣的内容，该过程基于用户、内容和其他辅助信息的语义知识。^[4]

常见的推荐场景有电商推荐，如亚马逊、阿里巴巴淘宝等；短视频/视频推荐，如抖音、快手、YouTube、腾讯视频等；新闻推荐，如今日头条、腾讯新闻等。音乐推荐，如 Netflix 等。不同的推荐场景具有不同的业务特点，如电商推荐关注下单率和购买金额，视频推荐关注视频停留时间，新闻推荐关注点击率以及新闻的时效性和准确性。可以说，推荐已经成为互联网发展的引擎，并且是目前互联网公司的核心业务。好的推荐系统会为公司带来上亿元的利润增长。

推荐的关键包含两方面内容：用户的偏好，内容的属性。这天然对应图表示学习中的 user 节点和 item 节点，即可认为通过增强 user 节点或 item 节点的代表来获取更佳的推荐效果。user 节点表示向量包含更多更准确 user 偏好的语义，item 节点表示向量包含更丰富商品内容的信息，在最后通过矩阵分解协同过滤模型就能得到更准确的推荐结果。

1.2 知识图谱与推荐

知识图谱是一类“特殊”的图：相较于 HIN，KG 包含的边类型更丰富，反应的深层特征/区别是 KG 注重节点之间的联通性，即目前 HIN 更加注重图结构即节点之间是否连接，而 KG 更注重两个节点之间以什么关系连接。因此，KG 的表示学习必须考虑边的类型和语义向量，这也为基于 KG 的推荐系统带来了优势和难度：优势在于可以将包含丰富语义信息的 KG 作为辅助信息来增强 user/item 节点，难度是目前 HIN 上的图神经网络模型难以适用于 KG。

除可以将 KG 作为辅助信息增强节点表示从而提升推荐效果这一优点外，基于 KG 的推荐还有个优势是“可解释性”。HIN 中一对 user 和 item 节点的点积大只能模糊的解释为 user 节点对 item 有较大偏好，而 KG 由于其数据本身的联通性，往往可以找到一对 user 和 item 节点之间的路径，从而从语义上解释

为什么向该 user 节点推荐该 item 节点，一个典型例子如图 1.1所示。

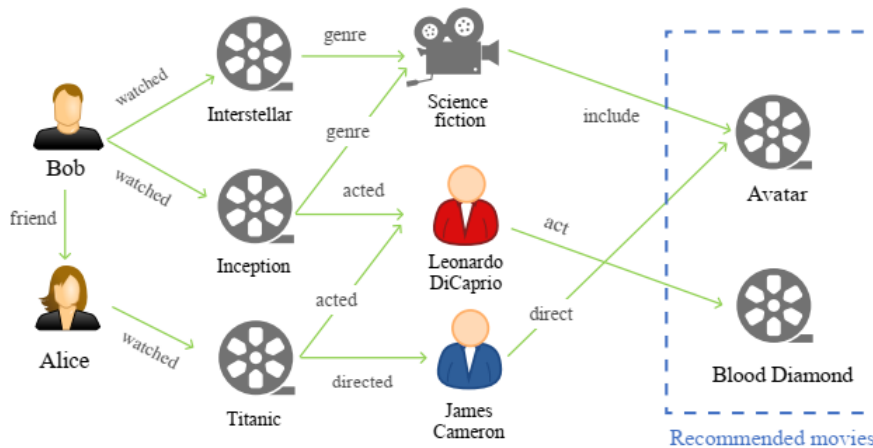


图 1.1 KG 推荐可解释示意图

第 2 节 基于知识图谱的推荐算法

这里我们按照模型本质分为 3 类。

1. 基于知识图谱嵌入 KGE 的方法
2. 基于知识图谱路径的方法
3. 基于图神经网络 GNN 的方法

2.1 基于 KGE 的算法

基于 KGE 的算法是一种自然而传统的方法，因此 KGE 本身就是 KG 的重要基础性工作，典型的 KGE 方法有 TransE^[1]，TransH^[15]，TransR^[7]，TransD^[6]，DisMult^[16] 等方法。在得到嵌入后，再把嵌入向量带入到推荐模块中，这一串联的过程是自然想到的一类方法。除此之外，近些年也有基于多任务学习而将 KGE 和 RS 的 loss 放在一起进行训练的方法。

KG 多由 item 及其属性节点构成，因此一个自然的想法就是利用 KG 增强 item 节点从而提升推荐效果。除此以外，也有将 user 节点纳入 KG 中，从而增强 user 和 item 节点。因此，基于 KGE 的 RS 算法可分为三类：

1. KG 只包含 item 节点的方法 (只增强 item 节点)
2. KG 同时包含 user 和 item 节点的方法
3. KGE 和 RS 多任务学习的方法

(1) KG 只包含 item 节点的方法

CKE^[17] 是 2016 年在 KDD 上发表的一篇用 KG 增强 item 节点的文章。其

item 节点包含四部分组成内容：1 由 item 知识图谱通过 TransR 得到的表示；2 由图像信息 CV 得到的表示；3 由文本 NLP 得到的表示；4 偏置项。最后将增强的 item 表示带入到 CF 框架中。CKE 是一篇经典的用 item 知识图谱增强 item 节点表示的文章。CKE 模型如图 2.1所示。

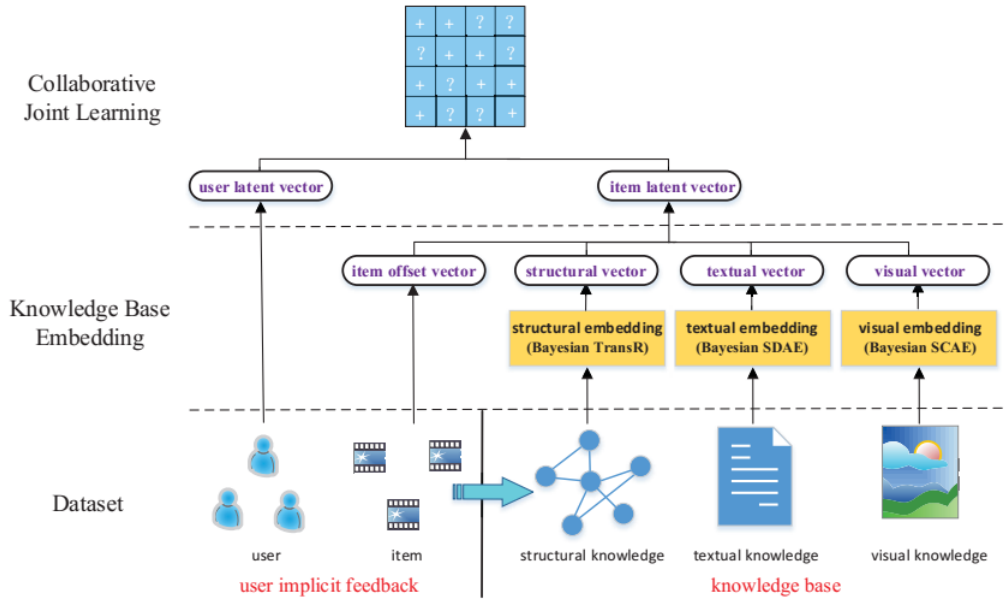


图 2.1 CKE 模型示意图

DKN^[11] 是 2018 年在 WWW 上发表的一篇用 KG 增强单词向量从而进行新闻推荐的工作。该算法可分为三步：1 首先构建单词的 item 知识图谱，采用了自定义规则，从而得到了 knowledge-level 的 word embedding，如图 2.2所示；2 再基于 KCNN 得到高质量的 sentence embedding；3 最后基于注意力机制将候选新闻的嵌入与用户浏览过新闻的嵌入转化为用户嵌入，再带入 CF 框架计算偏好程度。DKN 模型如图 2.3所示。

该技术路线下 item 知识图谱的核心作用就是作为辅助信息增强 item 表示，无论是 CKE 丰富了 item 表示的源数据种类/原始特征（除了图像和文本，增加了 KG），还是 DKN 里用 KG 增强基础层单词的表示再继而获得更好的句子表示，无论是并联还是串联的方式，都深刻地体现了利用 KG 增强 item 表示这一思想。

（2）KG 同时包含 user 和 item 节点的方法

KG 同时包含 user 和 item 节点的 KGE 推荐算法，与第一小类只包含 item 节点的方法无本质区别，只不过该类方法同时得到了 user 节点的嵌入，之后再

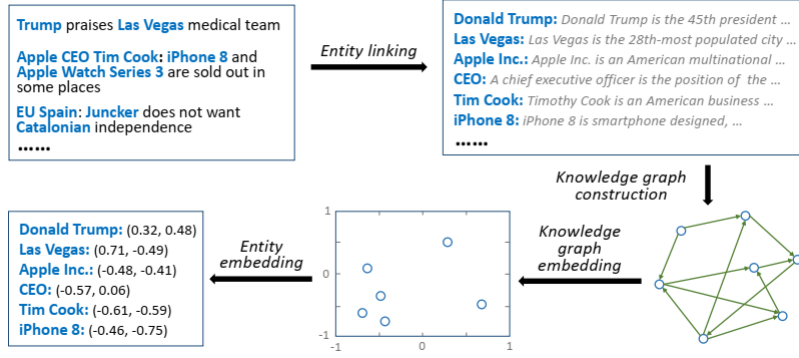


图 2.2 DKN 新闻单词知识图谱

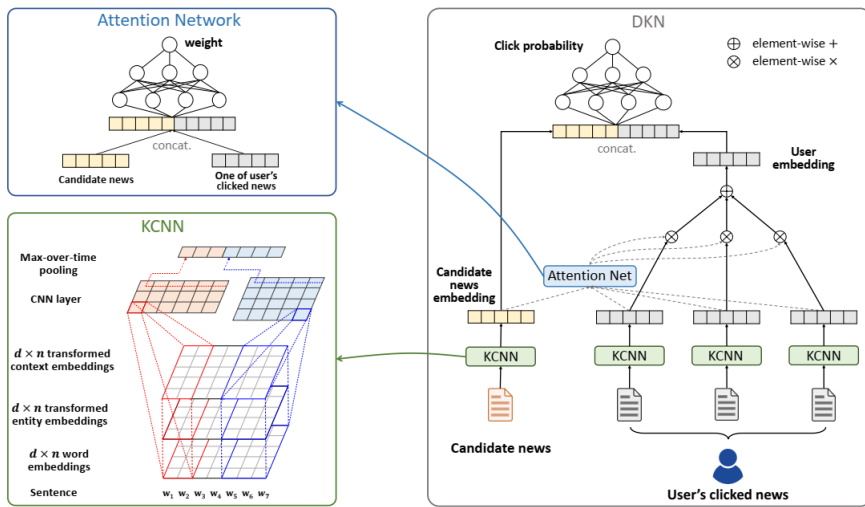


图 2.3 DKN 模型示意图

带入推荐框架进行排序。

CFKG 是 2018 年在 xiKiv 上的一篇典型 KGE 路线下同时包含 user 和 item 节点的文章。CFKG 构建了一个包含 user、item 和各类 user、item 属性的标准 KG，如图 2.4 所示。之后采用 KG 中常用的距离模型得到各节点表示 (即同时得到了 user 和 item 节点表示)。最后该文章与其他文章不同的是推荐模块没有采用 CF 框架，而是采用了另一种设计：将 user buy item 采用 KG 中三元组的距离模型思路，按照式 $distance(u_i + r_b u_y, v_j)$ 考虑给定 user、给定 interaction (推荐中常用 buy、click 等)，给定 item 的距离，距离越近认为该 user 越喜欢该 item。这一不同于大多数文章，采用了 KG 三元组距离模型思想的推荐偏好程度计算模块，值得借鉴。

SHINE^[10] 是 2017 年发表在 WSDM 上的一篇用于名人推荐的文章。该文

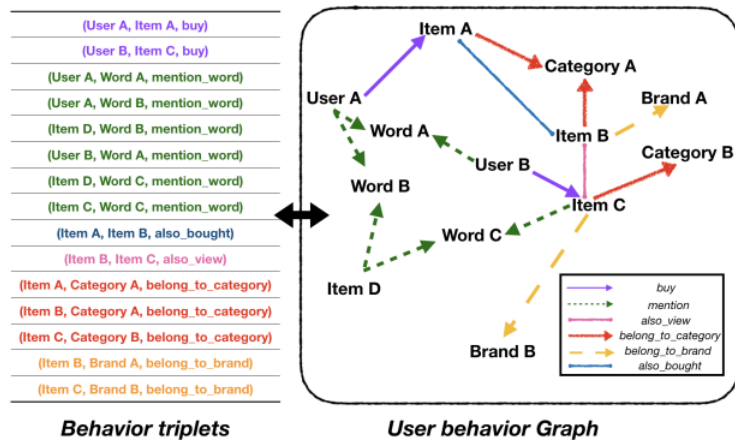


图 2.4 CFKG 构建的 KG

章旨在通过预测用户对名人的情绪来进行推荐。该文章一共构建了三类图，分别是情绪图、社交图、属性图，(这三个图其实更像是 HIN 而不是 KG)，这三幅图各自均包含 user 和 item 节点，分别嵌入得到三套表示，再集合起来带入协同过滤推荐框架。SHINE 其实更像是多图推荐 (即任务的输入包含多个种类的图，不是单一的图) 的典型算法。SHINE 模型示意图如图 2.5所示。

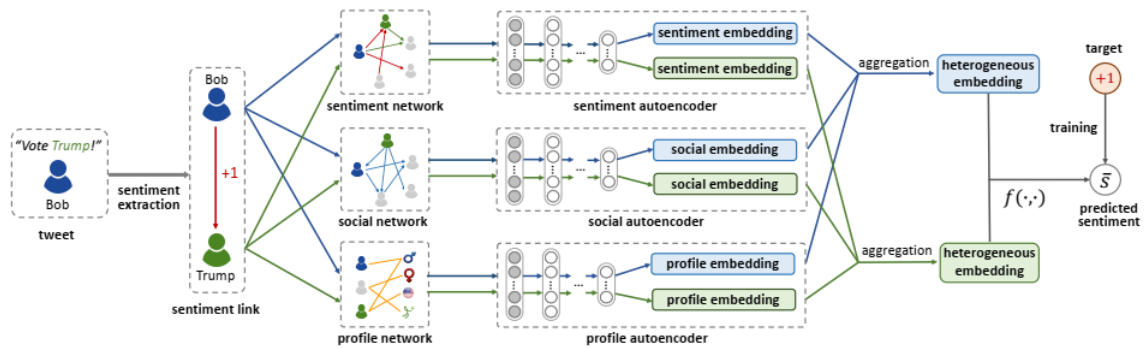


图 2.5 SHINE 模型示意图

(3) KGE 和 RS 多任务学习的方法

前两种方法都是先训练得到节点嵌入，再直接使用训练好的嵌入进行推荐预测。其实，这一设计从 loss 设计角度就存在一些缺陷，即没有对推荐效果设计 loss 进行训练。因此，一个自然想法是在 loss 中同时考虑 KGE 和 RS。但是值得注意的是，用于 KGE 的向量和用于 RS 的向量需要交互，即多任务学习不能割裂得各自计算 loss 最终粗暴的合在一起训练模型，而是在两个 loss 计算过程中考虑合理的交互机制，如何合理的交互是多任务学习的难点也是带来优势

的原因。

KTUP^[2] 是 2019 年发表在 WWW 上的一篇关于 KGE 和 RS 多任务学习的工作。该篇对 user 和 item 节点进行了两部分并列处理：第一部分按照 RS 任务，类似于 CFKG 中采用 KG 三元组距离模型来计算推荐偏好程度，据此得到 RS 的 loss L_{RS} 。第二部分按照 KGE 任务，采取距离模型计算包含 item 的知识图谱上节点嵌入并据此得到 KGE 的 loss L_{KGE} 。最后，采用多任务学习 $loss_{all} = L_{RS} + \gamma L_{KGE}$ 进行训练。同时，该篇采用了基于超平面投影的 TransH 方法来计算三元组中头实体、关系和尾实体在投影空间内的距离。KTUP 模型示意图如图 2.6 所示。

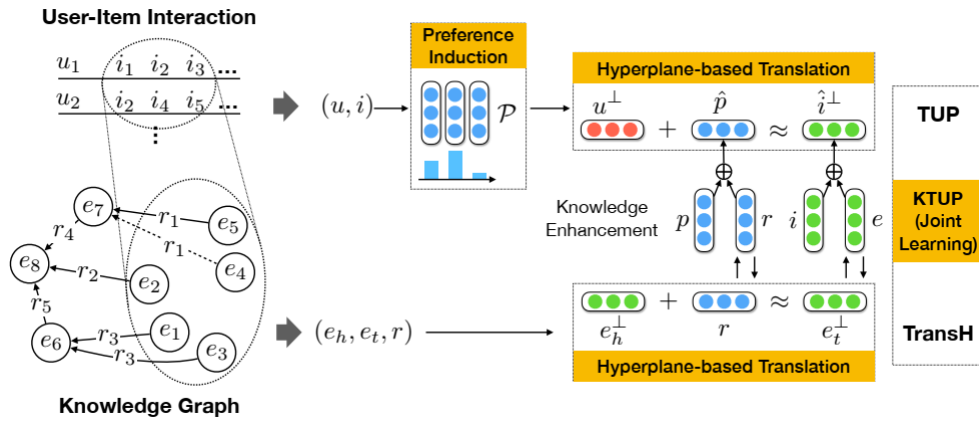


图 2.6 KTUP 模型示意图

MKR^[3] 是 2019 年发表在 WWW 上与 KTUP 类似技术途径的工作。但是个人认为 MKR 的方法更加通用。MKR 也是分两部分计算 RS 和 KGE 的 loss，但是对原始表示向量经过了多层全连接层，这两部分的全连接层之间存在交互，交互方式为 item 向量和其对应的实体嵌入共同作为输入，经过权重和偏置项得到新的 item 向量和实体嵌入；即该交互利用了同一个 item 的两套表示之间的关系。MKR 模型示意图如图 2.6 所示。

2.2 基于知识图谱路径的方法

知识图谱和 HIN 都是图结构化数据，从定义上说 KG 属于 HIN，但目前 HIN 和 KG 仍有较大差别。其本质差别就是 KG 更注重节点间的联通性，因为 KG 更注重边的类型和内容/表示。因此，在 HIN 上分析两个节点之间的关系可能无法通过他们之间的路径得到，而 KG 则大多可以通过分析两个节点间的路径得到他们交互的原因和关系。总的来说，就是 KG 相较于 HIN 其节点间，尤其是多跳路径，具有丰富语义和有效信息，需要被合理利用。因此，此类方法旨

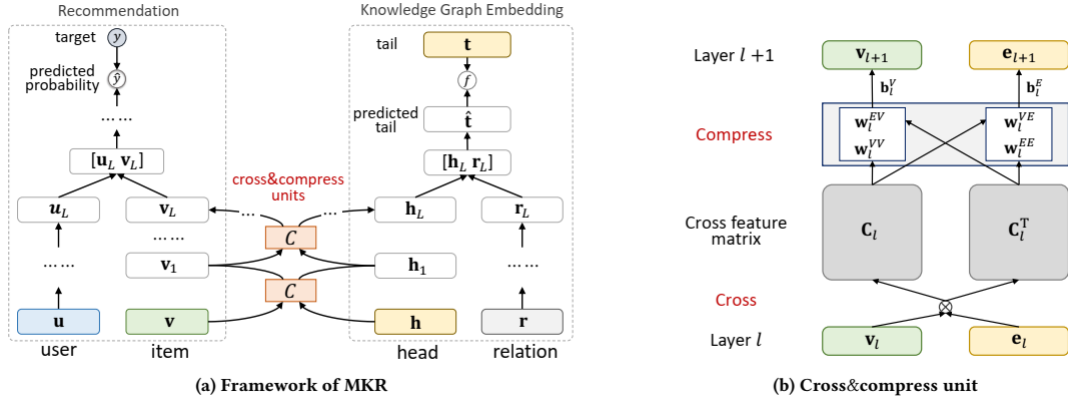


图 2.7 MKR 模型示意图

在利用节点间路径从而得到更好地推荐效果，这是与第一类方法本质区别的地方。

MCRec^[5] 是发表在 2018 年 KDD 上的一篇利用了元路径编码的工作。MCRec 首先对每个元路径采样多个实例，再对每个实例基于 CNN 进行编码，最后采取最大池化得到每个元路径的编码。值得注意的是，每个元路径的编码与具体节点无关，即 MCRec 认为每条元路径代表一个推荐理由，每个理由不随具体的用户或商品而改变。之后在计算推荐匹配函数时，对每一对 user-item 之间不同的元路径采取注意力机制计算最终打分。MCRec 模型示意图如图 2.8 所示。

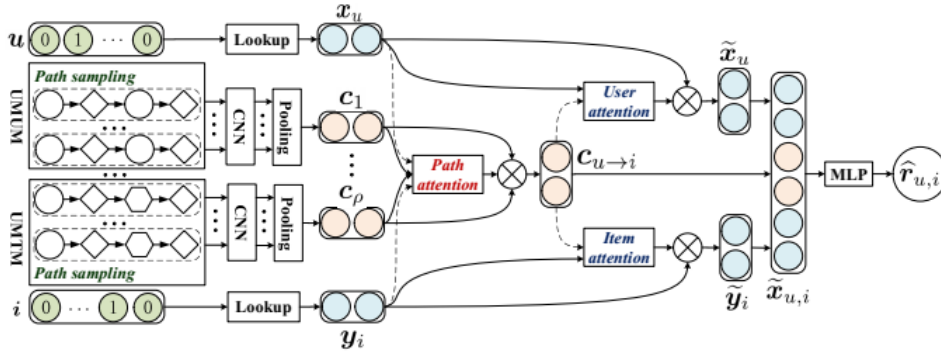


图 2.8 MCRec 模型示意图

RKGE^[9] 是发表在 2018 年 RecSys 上利用节点对间路径进行推荐的工作。RKGE 与 MCRec 相同的是都对节点对间路径进行了编码，不同的是 MCRec 认为每条元路径编码固定与节点内容无关，RKGE 认为每条路径的编码都不一样即与其路径上的节点相关。RKGE 首先采样指定 user-item 节点对之间的所有路径，对每条路径通过 RNN 进行编码 (最后一个隐藏单元作为整条路径编码)，最

后采取平均池化对同一对节点间所有路径编码处理为一个编码，代表该 user 对该 item 的喜好程度。RKGE 尽管模型简单，但是突出体现了本节开始提到的一点：利用 KG 相较于 HIN 其节点间具有丰富语义和有效信息的路径这一特点得到更好地推荐效果。RKGE 模型示意图如图 2.9所示。

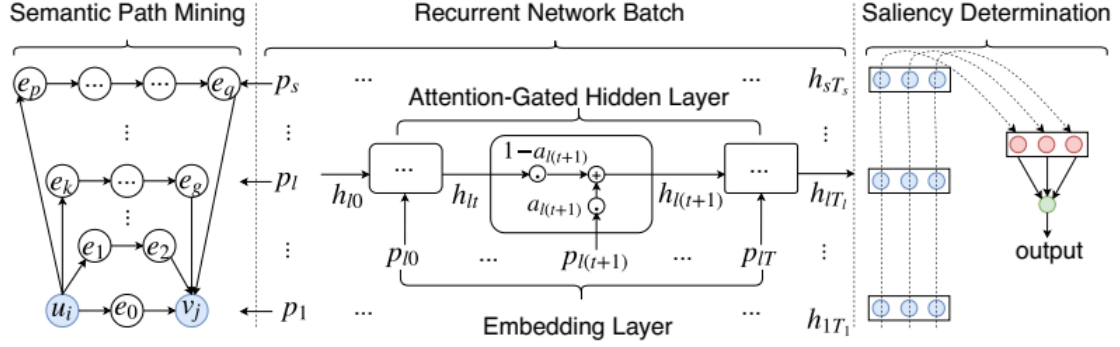


图 2.9 RKGE 模型示意图

2.3 基于 GNN 的算法

前两条技术路线，一是利用了传统 KGE 技术，二是使用了 kG 中对路径的语义学习表示，这两个本质思想都是 KG 推荐的经典和有效途径。但，近些年兴起的 GNN 技术，为 KG 带来了新的发展方向，尤其是基于 GNN 的 KG 推荐模型。但是目前 GNN 更多研究方向在于中心节点如何聚合周围节点 (无论是是否使用元路径)，而 KG 的突出特点是边种类和内容不可忽视，因此 GNN 路线下的 KG 推荐的突出和首要问题就是如何设计适用于 KG 的 GNN 框架。除此之外，还有图谱融合和预处理的过程，即知识图谱不仅包含有利于推荐的语义信息，还包含与所关系的推荐问题无关的冗余信息甚至是“噪声” (这里噪声信息不是指数据错误或包含误差，而是指在 GNN 框架下不利于节点表示的聚合信息)。因此，尽管 KG 增强 RS 是很有优势的一条技术路线，但是也是一条较有难度的技术路线。这里我们介绍四个经典模型

KGCN^[13] 是第一篇将 KG 用于 RS 的文章，该文章最大的贡献是提出了如何将 GNN 应用于 KG，具体地说，就是在对中心节点聚合时，将边信息纳入对相邻节点做 attention 时的系数中，其 motivation 是不同的用户对相同的边有不同的偏好，例如用户 A 更注重价格属性，用户 B 更注重品牌属性。KGCN 核心公式如式 1 所示，其模型示意图如图 2.10 所示。

$$\begin{aligned}
\alpha_{r_{e_i, e_j}}^u &= u^T r_{e_i, e_j} \\
\beta_{r_{e_i, e_j}}^u &= \frac{\exp(\alpha_{r_{e_i, e_j}}^u)}{\sum_{k \in N(e_i)} \exp(\alpha_{r_{e_i, e_k}}^u)} \\
v_{N(v)}^u &= \sum_{e \in N(v)} \tilde{\pi}_{r_v, e}^u e
\end{aligned} \tag{1}$$

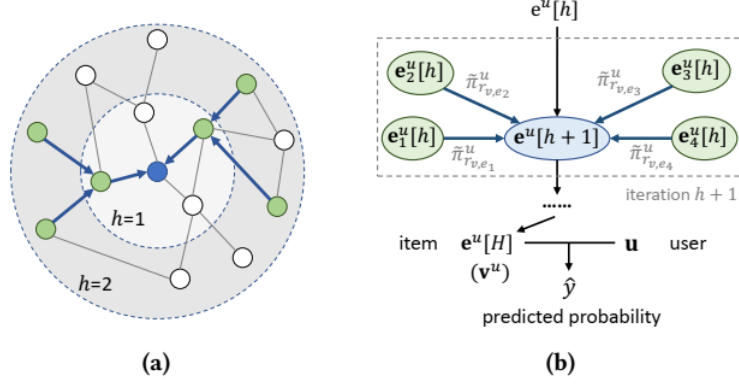


图 2.10 KGCN 示意图

KGCN-LS^[12] 与 KGCN 是同一批作者，投在了不同的 A 会上，这两篇核心思想一致，KGCN-LS 同样采用 KGCN 中在边中纳入用户偏好，KGCN-LS 采用的 GNN 框架是 GCN，但是在连接矩阵中纳入了可训练的向量，同时针对该连接矩阵在 loss 中增加了一项正则项：类似于标签传播的 idea，基于连接矩阵中边的权值大小，直接将已有标签的节点和标签向图上其余节点传播，其传播规则是能量函数最小。KGCN-LS 核心公式如式 2 所示，其模型示意图如图 2.11 所示。

$$\begin{aligned}
A_u^{ij} &= u^T r_{e_i, e_j} \\
l_u^*(e_i) &= \frac{1}{D_u^{ii}} \sum_{e_j \in \sigma} A_u^{ij} l_u^*(e_j)
\end{aligned} \tag{2}$$

KGAT^[14] 是另外一种将 GNN 应用于 KG 的工作。KGAT 对所有节点不加区分，将 user 节点纳入知识图谱中。聚合时，邻居节点的 attention 系数由头实体、关系和尾实体共同组成。KGAT 的 motivation 是这样的模型可以捕捉更高阶的连接信息，从而更充分地利用知识图谱里的语义信息。笔者个人认为，尽管在推荐中，我们更关心图上的 user 和 item 节点，对于其他节点并不关心他们的表示，因此在模型设计中，可以有区别得或者有针对性得去增强 user 和 item 节

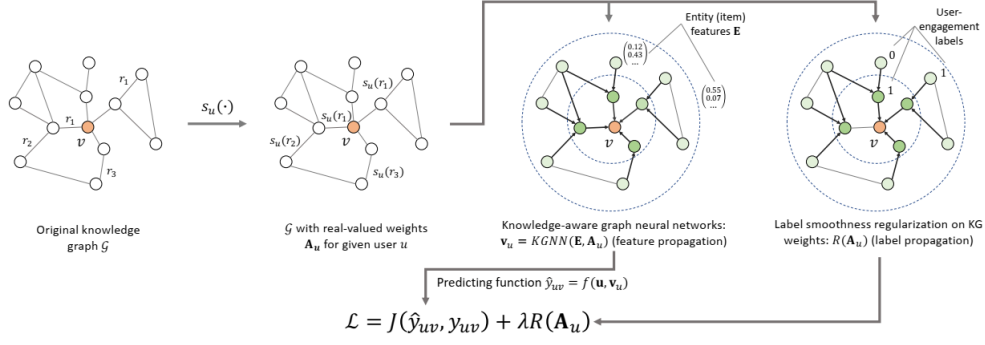


图 2.11 KGAT-LS 示意图

点。但是，KGAT 提供了一个不加区分的节点表示方法，这一技术思路值得借鉴。KGAT 的核心公式如式 3 所示，其模型示意图如图 2.12 所示。

$$\pi(h, r, t) = (W_r e_t)^T \tan(W_r e_h + e_r)$$

$$e_{N_h}^u = \sum_{(h,r,t) \in N_h} \pi(h, r, t) e_t \quad (3)$$

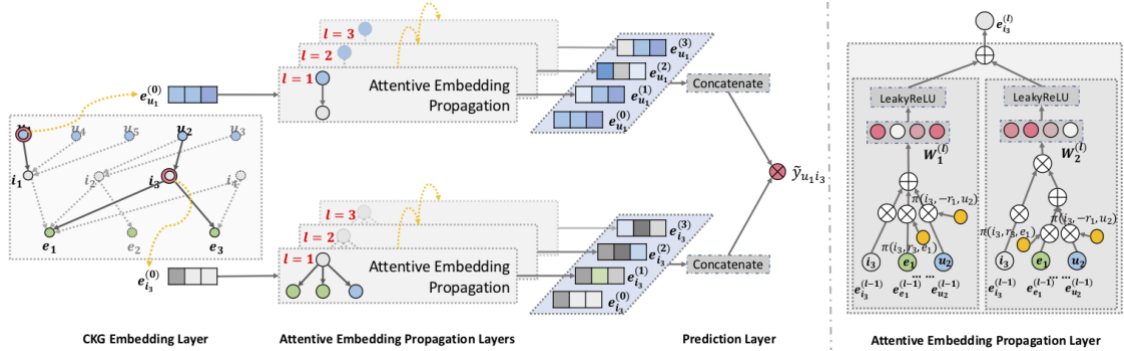


图 2.12 KGAT 示意图

AKGE^[8] 是一个对原始 KG 简化和抽取子图的经典模型。AKGE 在原始 KG 中，首先按照距离最短原则 (此处距离不是常规的跳数，而是表示向量欧式距离的和最短，此时欧氏距离代表语义接近程度) 抽取出目标 user 节点和 item 几点之间的路径。在得到子图之后，用于 KG 的 GNN 与 KGAT 类似，也是不加区分节点的聚合模型，此处聚合时采用节点类型和边类型增强的方法 (每个节点先采用节点类型增强，聚合时再根据聚合的中心节点采用边类型增强)。AKGE 是一篇经典之作，其创新点包含如何抽取子图和如何将 GNN 应用于 KG 两个方面。AKGE 的核心公式较多，此处不再赘述，详见原文^[8]，其模型示意图如图

2.13所示。

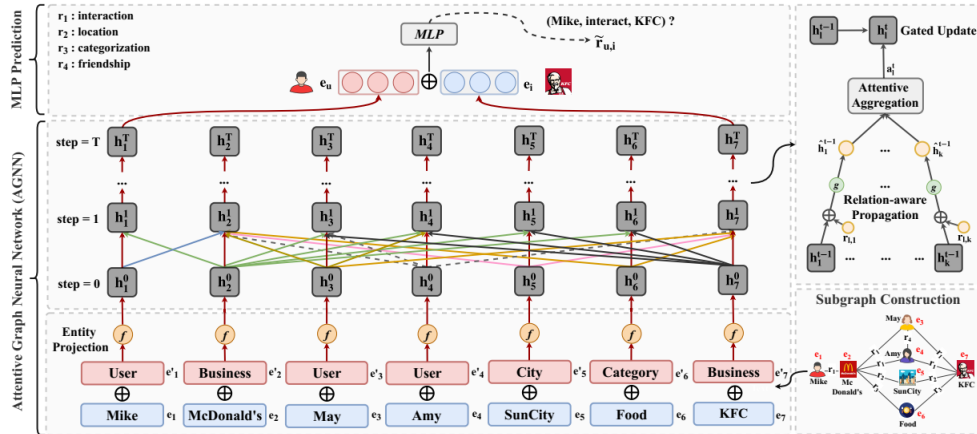


图 2.13 AKGE 示意图

第 3 节 小结

基于 KGE 的推荐算法，无论是 KG 中只包含 item 节点还是同时包含 user 和 item 节点，都是通过 KGE 先得到增强的节点嵌入，然后带入到 RS 框架中；有时在设计 loss 时会将 KGE 和 RS 的 loss 均考虑即多任务学习。

基于 KG 路径的推荐算法，旨在利用以下特点：KG 不同于其他图结构化数据，KG 更注重节点间联通性，即 KG 可以通过两个节点间的路径得到他们交互的原因和关系，而其他图可能不具备此特征。因此，此类方法旨在利用节点间路径的丰富语义和有效信息从而得到更好地推荐效果。

基于 GNN 的推荐算法，效果有显著提升，但存在两个突出技术困难：如何设计适用于含有丰富边类型 KG 的 GNN 模型，如何在丰富甚至冗余或噪声 (这里噪声指与人物无关，不是信息错误) 的语义信息中利用与任务相关的那部分信息。

本次大作业旨在对最近本人对基于知识图谱的推荐系统的调研做一个总结，同时完成 2020 年人工智能前沿讲座选做大作业中刘康老师布置的作业。这篇综述根据 KG 推荐方法的技术不同，较为系统地总结了近些年最新的研究成果，并给出了个人心得体会。本人研究方向为“图数据挖掘和基于图的推荐系统”，欢迎交流讨论。(huaizepeng2020@ia.ac.cn)

第 4 节 参考文献

- [1] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko. Translating embeddings for modeling multirelational data. *NIPS*, 2013.
- [2] Y. Cao, X. Wang, X. He, Z. Hu, and T.-S. Chua. Unifying knowledge graph learning and recommendation: Towards a better understanding of user preferences. *WWW*, 2019.
- [3] Y Cao, W Xiang, X He, Z Hu, and Tat-Seng Chua. Unifying knowledge graph learning and recommendation: Towards a better understanding of user preferences. *WWW*, 2019.
- [4] Kim Falk. Practical recommender systems.
- [5] Binbin Hu, Chuan Shi, Wayne Xin Zhao, and Philip S. Yu. Leveraging meta-path based context for top- n recommendation with a neural co-attention model. In *the 24th ACM SIGKDD International Conference*, 2018.
- [6] G. Ji, S. He, L. Xu, K. Liu, and J. Zhao. Knowledge graph embedding via dynamic mapping matrix. *ACL*, 2015.
- [7] Y. Lin, Z. Liu, M. Sun, Y. Liu, and X. Zhu. Learning entity and relation embeddings for knowledge graph completion. *AAAI*, 2015.
- [8] Xiao Sha, Zhu Sun, and Jie Zhang. Attentive knowledge graph embedding for personalized recommendation. *arXiv*, 2020.
- [9] Z Sun, J Yang, J Zhang, A Bozzon, LK Huang, and C Xu. Recurrent knowledge graph embedding for effective recommendation. In *RecSys*, 2018.
- [10] H Wang, F Zhang, M Hou, X Xie, M Guo, and Q Liu. Shine: Signed heterogeneous information network embedding for sentiment link prediction. *WSDM*, 2017.
- [11] H. Wang, F. Zhang, X. Xie, and M. Guo. Dkn: Deep knowledge-aware network for news recommendation. *WWW*, 2018.

- [12] Hongwei Wang, Fuzheng Zhang, Mengdi Zhang, Jure Leskovec, Miao Zhao, Wenjie Li, and Zhongyuan Wang. Knowledge-aware graph neural networks with label smoothness regularization for recommender systems. 2019.
- [13] Hongwei Wang, Miao Zhao, Xing Xie, Wenjie Li, and Minyi Guo. Knowledge graph convolutional networks for recommender systems. *WWW*, 2019.
- [14] Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat Seng Chua. Kgat: Knowledge graph attention network for recommendation. 2019.
- [15] Z. Wang, J. Zhang, J. Feng, and Z. Chen. Knowledge graph embedding by translating on hyperplanes. *AAAI*, 2014.
- [16] Bishan Yang, Wen Tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. Embedding entities and relations for learning and inference in knowledge bases. *arXiv*, 2014.
- [17] F. Zhang, N. J. Yuan, X. Xie D. Lian, and W.-Y. Ma. Collaborative knowledge base embedding for recommender systems. *KDD*, 2016.