《metapath2vec: Scalable Representation Learning for Heterogeneous Networks》

《metapath2vec:异构网络的可扩展表示学习》

**摘要：**

1. **本文的背景：**研究者们发现基于同质图的图表征学习算法，例如Deepwalk，Node2Vec等并不能很好地直接应用到点和边有多个类型的异质图上。而这主要是由于这些同质图算法并不考虑节点的类型，导致它们往往有如下两方面的缺点：1）容易偏向于出现频率高的节点类型；2）容易偏向于连接相对集中的节点。因此，本文旨在解决上述网络异构所带来的特殊挑战，从而进一步扩展图表征学习的应用范围。
2. **本文的贡献：**提出了一种专门处理异质图的图嵌入学习框架——metapath2vec，以及其改进版本metapath2vec++，有效保持了网络结构在语义上的相关性。
3. **主要创新点：**首次正式定义了Heterogeneous Network上的表示学习过程；提出了一种基于元路径的随机游走方法以适用于异构图网络。
4. **实验结果：**本文通过大量的实验证明了本文所提出的方法在各种异构网络挖掘任务中的有效性和可扩展性，如节点分类（相对于基准提高了35%—319%）和节点聚类（相对于基准提高了13%—16%）。此外，还证明了metapath2vec和metapath2vec++能自动发现异构网络中不同类型节点之间的内部语义关系，而现有工作是无法发现的。

**问题定义：**

异构图的定义：

给定一个图，对于其中每个节点和边都有对应的映射函数：，。其中，和分别表示节点和边的类型集合，且。即，图网络的节点类型与边类型的总量大于2。

异构图表示学习的定义：

给定一个异构图，其任务是学习到所有节点的嵌入表示，其中，而嵌入表示能有效捕捉到不同类型节点的结构和语义关系。

**注意：**虽然异构图上的节点类型不同，但是不同类型的节点的嵌入向量会映射到同一个特征空间中，即不同类型的节点的嵌入向量的维度都是。此外，由于网络异构性的存在，传统的基于同构网络的节点嵌入表征方法很难有效地直接应用在异构网络上。

**异质skip-gram：**

传统的skip-gram模型：

给定一个网络，其目标是使网络的局部结构概率最大化，即：

其中，代表节点的上下文节点，而这种节点的定义方式有多种，比如中利用随机游走得到的上下文节点。而代表在已知节点的情况下，上下文节点存在的概率，这种概率通常用向量内积来实现：

而为了将异构网络的结构引入到上述的skip-gram方法中，本文提出了异质skip-gram的概念，即在给定节点的情况下，最大化其异质上下文节点出现的概率，即：

其中，代表节点的第种类型的上下文节点；代表在已知节点𝑣的情况下，上下文节点存在的概率，而这种概率在本文也是用向量内积+𝑠𝑜𝑓𝑡𝑚𝑎𝑥来实现的：

此外，与传统的skip-gram中一致，为了解决每次更新时都要对所有节点进行计算的问题，异质skip-gram也引入了负采样策略。给定负样本数，则优化目标可进一步表示为：

其中，代表负采样中样本的预定义分布；激活函数；代表从负样本集中抽取的负样本。

**基于元路径的随机游走：**

**结论：**

**对本文的感悟：**