Graph Embedding

陈浩

2020.04

Item2vec

soft	腾讯微云;必胜客;指划修图 Snapseed;动态壁纸选择器;微博;
	功能相机;微信;计算器;下载内容 夏新定制版;小米百变锁屏(MiLocker);京东
game	镜界;极品飞车最高通缉 OL;逗比人生;我的世界;火影战记;车祸英雄;穿越火线：枪战王者
	模拟山羊;我功夫特牛;指尖战车;我的世界;剑与远征

Item2vec是将word2vec的方法迁移到了推荐系统中

数据：将items视为word，用户的行为序列视为一个集合，item间的共现为正样本，并按照item的频率进行负采样，缺点是：忽略了user行为的序列信息，没有对用户对item的喜欢程度进行建模

模型：

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{-m\le j\le m,\, j\ne 0}\log p(w_{t+j}\mid w_t;\theta)$$

$$p(o\mid c)=\frac{\exp(u_o^T v_c)}{\sum_{w\in|T|}\exp(u_w^T v_c)}$$

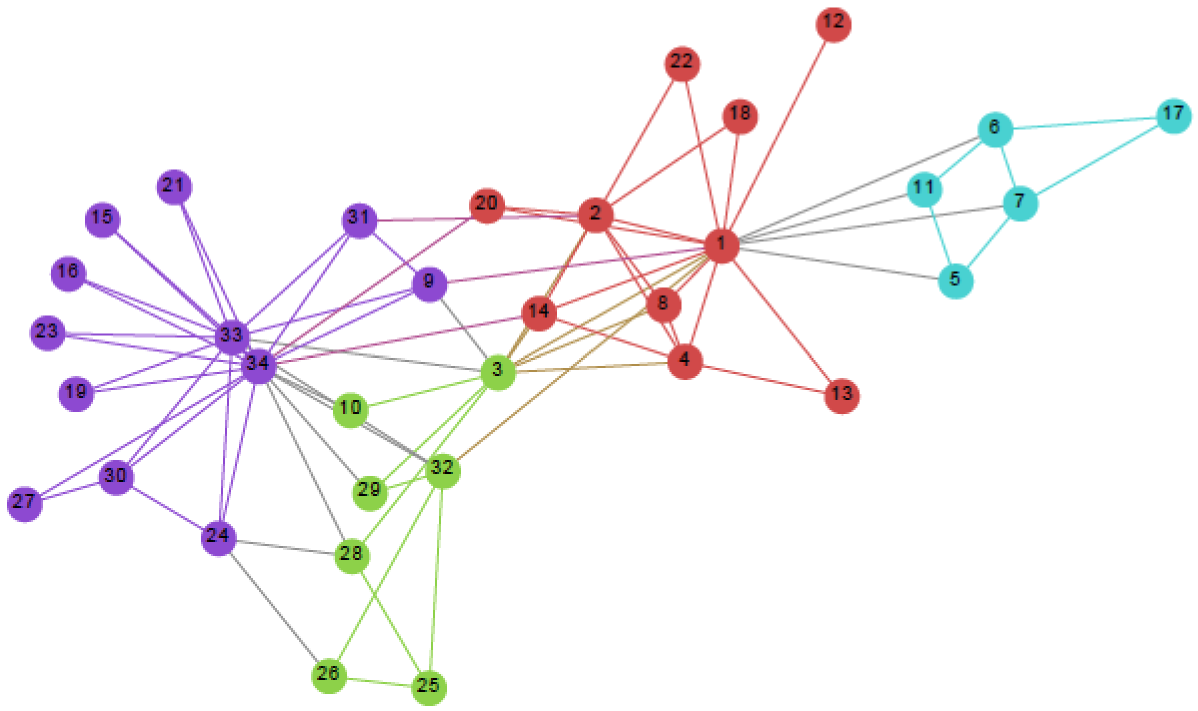
word2vec

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{j\ne t}\log p(w_j\mid w_t;\theta)$$

$$p(o\mid c)=\frac{\exp(u_o^T v_c)}{\sum_{w\in|T|}\exp(u_w^T v_c)}$$

item2vec

DeepWalk



deepwalk思想类似word2vec，用图中节点与节点的共现关系学习节点的向量表示

空间结构 => 序列结构

数据：deepwalk使用随机游走的方式在图中进行采样，是一种可以重复访问节点的深度优先遍历算法，在给定当前节点的情况下，从该节点的邻居节点中随机采样作为下一个节点，重复此过程，直到序列长度满足预设的条件

模型：采样结束后，得到节点的序列结构数据，利用word2vec的思想训练模型，并通过Skip-gram结合负采样的方式优化模型，进而学习图中节点的向量表示

LINE

一阶相似度用于描述图中的局部相似度
对于节点 v_i 和节点 v_j 边权 w_{ij} 即为其相似度

联合概率：

$$p_1(v_i, v_j) = \frac{1}{1 + \exp(-u_i^T \cdot u_j)}$$

经验分布：

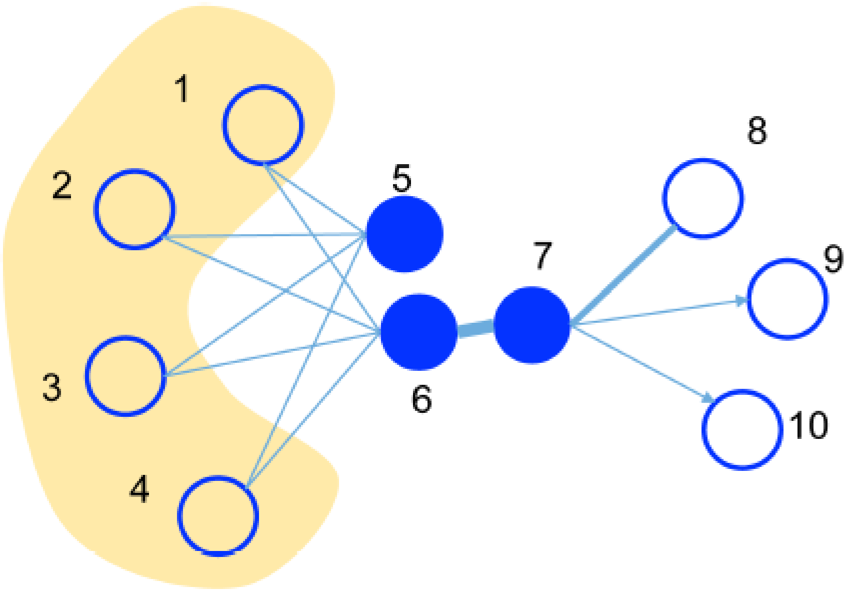
$$\hat{p}_1(v_i, v_j) = \frac{w_{ij}}{W}; \quad W = \sum_{(i,j)\in E} w_{ij}$$

优化目标：

$$O_1 = d(\hat{p}_1(\cdot,\cdot), p_1(\cdot,\cdot))$$

$$= -\sum_{(i,j)\in E} \hat{p}_1(v_i, v_j) \log \frac{p_1(v_i, v_j)}{\hat{p}_1(v_i, v_j)}$$

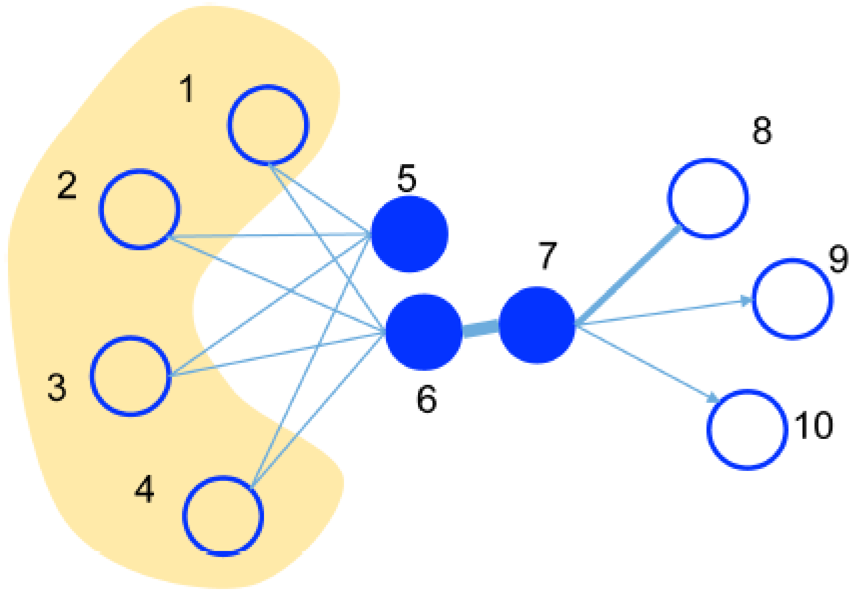
$$\Leftrightarrow -\sum_{(i,j)\in E} w_{ij} \log p_1(v_i, v_j)$$



节点6和节点7直接相连，因此这两个节点有较高的一阶相似度

LINE

二阶相似度用于描述图中的有相同 context 节点的相似度，每个节点有两个embedding，一个是节点本身的向量表示 u_i，另一个是作为 context 的向量表示u'_i



节点5和节点6有相同的context，因此这两个节点有较高的二阶相似度

对于一条边 (i, j) 给定节点 i 则产生节点 j 的概率为：

$$p_2(v_j \mid v_i) = \frac{\exp(u'_j \cdot u_i)}{\sum_{k=1}^{|V|} \exp(u'_k \cdot u_i)}$$

经验分布：

$$\hat{p}_2(v_i \mid v_j) = \frac{w_{ij}}{d_i}$$

d_i 为节点 i 的出度

优化目标：

$$O_2 = \sum_{i \in V} \lambda_i d(\hat{p}_2(\cdot \mid v_i), p_2(\cdot \mid v_i))$$

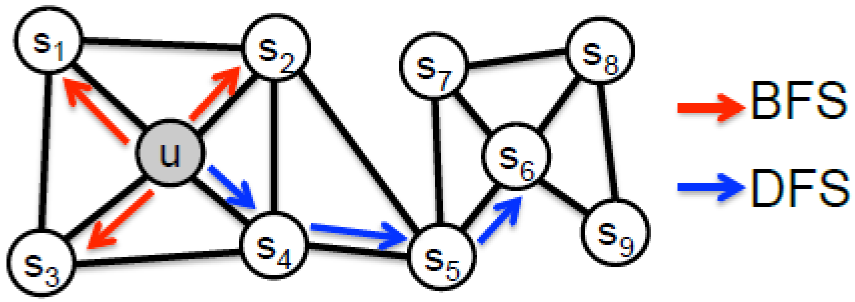
$$\Leftrightarrow -\sum_{(i,j) \in E} w_{ij} \log p_2(v_i \mid v_j)$$

λ_i 为节点节点的重要因子，可以用节点的出度或者 PageRank 等方法获取得到，这里取其为节点的出度

Node2vec

node2vec是对deepwalk的一种拓展，相较deepwalk的深度优先策略，node2vec兼顾了深度优先和广度优先的策略

数据：有偏的随机游走策略，将空间数据转化为序列数据



给定节点 v ,访问下一个节点 x 的概率为：

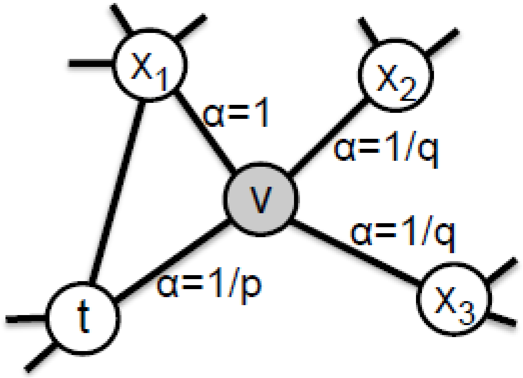
$$p(c_t = x \mid c_{t-1} = v) = \begin{cases} \dfrac{\pi_{vx}}{Z} & (v,x) \in E \\ 0 & otherwise \end{cases}$$

Z 为归一化常数

node2vec利用两个超参数 p和 q 来控制随机游走的策略，假设随机游走经过边(t,v)到达节点 x 则 $\pi_{vx} = \alpha_{pq}(t,x) \cdot w_{vx}$

$$\alpha_{pq} = \begin{cases} \dfrac{1}{p} & if \quad d_{tx} = 0 \\ 1 & if \quad d_{tx} = 1 \\ \dfrac{1}{q} & if \quad d_{tx} = 2 \end{cases}$$

d_{tx} 为节点 t 到节点 x 之间的最短路径距离



p 控制重复访问刚刚访问过的节点的概率
q 控制偏向BFS或者偏向DFS的概率

Node2vec

模型：
设$f(u)$为节点 u 的向量映射函数，定义$Ns(u)$为通过策略 S 采样得到的节点 u 的近邻点集合，node2vec通过最大化节点的近邻定点的概率即：

$$\max_f \sum_{u\in V} \log p(Ns(u) \mid f(u))$$

为了使上式可解，提出两个假设：
1. 独立性假设：给定节点 u,则与其近邻的节点出现的概率与其近邻的其他节点无关，则：

$$p(Ns(u) \mid f(u)) = \prod_{n_i \in N_s(u)} p(n_i \mid f(u))$$

2. 特征空间对称假设：相较LINE中的两个embedding，这里一个节点作为自身和作为context共享embedding向量则条件概率公式表示为：

$$p(n_i \mid f(u)) = \frac{\exp(f(n_i)\cdot f(u))}{\sum_{v\in V}\exp(f(v)\cdot f(u))}$$

最终的优化形式为：

$$\max_f \sum_{u\in V}[-\log Z_u + \sum_{n_1\in N_s(u)} f(n_i)\cdot f(u)] \qquad \text{其中 } Z_u = \sum_{v\in V}\exp(f(u)\cdot f(v))$$

Ali三种方式

BGE：Base Graph Embedding (基于概率随机游走)

GES: Graph Embedding with Sid Information

假设每个节点 v有 n 种不同的side information: $W_v^0, \ldots, \ldots W_v^n$

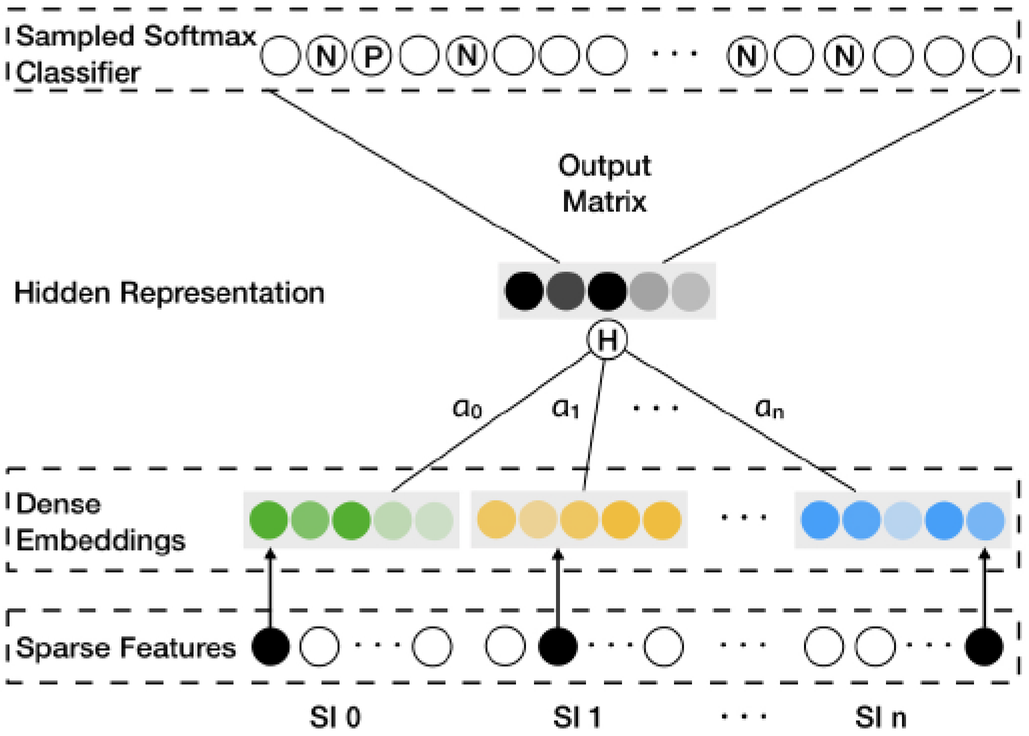
average-pooling operation:

$$H_v = \frac{1}{n+1}\sum_{s=0}^{n} W_v^s$$

EGES: Enhanced Graph Embedding with Side Information

weighted average combining:

$$H_v = \frac{\sum_{j=0}^{n} e^{a_v^j} W_v^j}{\sum_{j=0}^{n} e^{a_v^j}}$$



模型:

$$L(v,u,y) = -\left[y\log(\sigma(H_v^T Z_u)) + (1-y)\log(1-\sigma(H_v^T Z_u)) \right]$$

SDNE

设邻接矩阵：$S=\{s_1,\dots.s_n\}$
若取 $x_i=s_i$ 则：

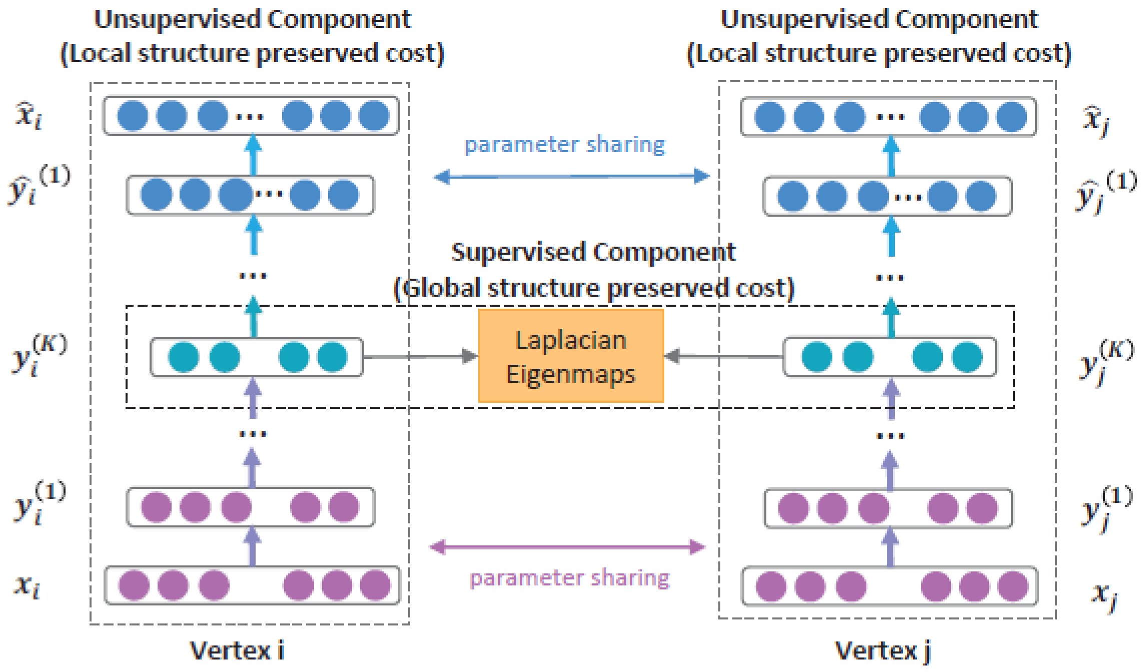
$$y_i^{(1)}=\sigma(W^{(1)}x_i+b^{(1)})$$
$$y_i^{(k)}=\sigma(W^{(k)}y_i^{(k-1)}+b^{(k)}),k=2,\dots,K$$

一阶损失：

$$L_{1st}=\sum_{i,j=1}^{n}s_{ij}\left\|y_i^{(K)}-y_j^{(K)}\right\|_2^2$$

二级损失：

$$L_{2nd}=\sum_{i=1}^{n}\left\|(\hat{x}_i-x_i)\odot b_i\right\|_2^2$$
$$=\left\|(\hat{X}-X)\odot B\right\|_F^2 \quad if\ \ s_{i,j}=0,b_{i,j}=1\ else\ b_{i,j}=\beta>1$$



总损失：$L=L_{2nd}+\alpha L_{1st}+\beta L_{reg}$

$$=\left\|(\hat{X}-X)\odot B\right\|_F^2+\alpha\sum_{i,j=1}^{n}s_{ij}\left\|y_i^{(K)}-y_j^{(K)}\right\|_2^2+\beta\frac{1}{2}\sum_{k=1}^{K}(\left\|W^{(k)}\right\|_F^2+\left\|\hat{W}^{(k)}\right\|_F^2)$$

谢谢！