

Discriminative Deep Random Walk for Network Classification

用于网络分类的判别式深度随机游走

ACL 2016

Juzheng Li, Jun Zhu, Bo Zhang

Dept. of Comp. Sci. & Tech., State Key Lab of Intell. Tech. & Sys. Tsinghua University,

实验背景

分类是自然语言处理工作中的一个重要任务，之前的工作着重于发展对各种各样的网络分析任务建立静态模型，将网络分类也常用于内容推荐，链路预测和异常检测，而静态模型的一个普遍的挑战就是网络的稀疏性。

解决这个问题的一个有效的策略就是在网络中为实体学习潜在空间表示，例如DeepWalk。如果实体之间有较近的关系，那么他们之间很有可能会分享相同的爱好或者属于同一小组。这对网络分类有很重要的作用

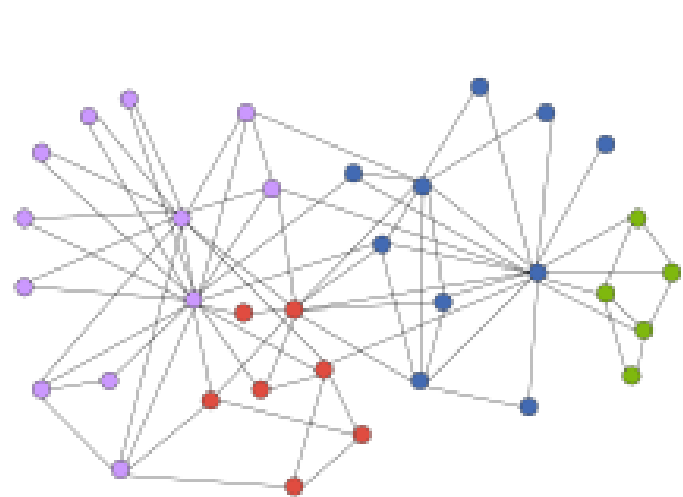
相关工作

在**关联网络**分类中，DeepWalk可能并不理想，因为它缺乏优化目标任务的机制。在这篇论文中，作者提出了一个用于关系网络分类的新方法——**判别式**深度随机游走（DDRW）。

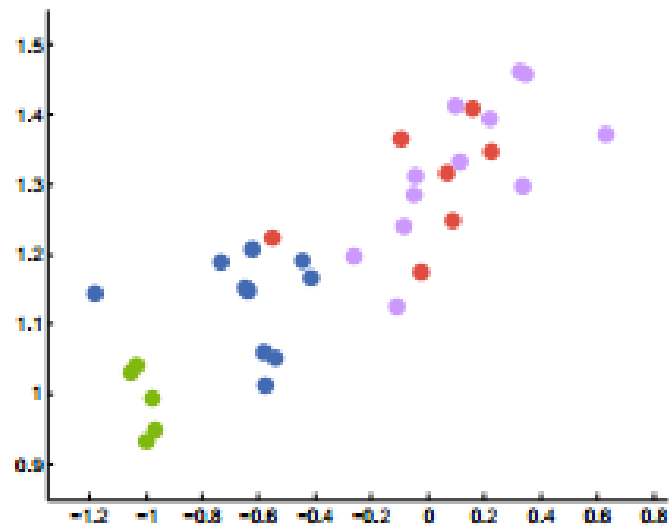
DDRW通过联合优化**分类目标函数**和**实体嵌入到潜在空间的目标函数**，扩展了DeepWalk.

文章使用Karate Graph，将网络嵌入到二维的真实空间，展示不同实验结果

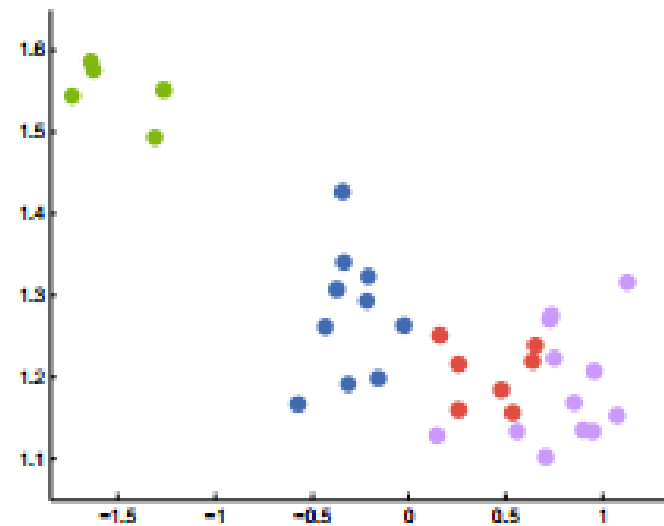
Karate Graph: 空手道俱乐部网络，该网络是社会网络分析领域中的经典数据集，20世纪一名社会学家用2年的时间观察美国一所大学空手道俱乐部34名成员间的社会关系，该网络中共有34个顶点，两个顶点之间有一条边则意味着相应的两个成员之间至少是交往频繁的朋友关系



(a) Karate Graph



(b) DeepWalk Embedding



(c) DDRW Embedding

论文思路

在这个连接学习平台中，DDRM设法去学习潜在表示与标签的强烈关系，很容易找到类之间的分离边界，而在原来网络中连接的actors在潜在社会空间中仍然关系紧密。这种结合特定任务和表示目标的想法在其他领域也被广泛探讨。

从技术层面来讲，为了从拓扑结构中捕获信息，我们采取和DeepWalk相似的想法：首先在原网络中利用截断随机游走获取actors的序列，然后建立一个从actors到潜在空间的Word2Vec的语言模型，为了将监督信号纳入网络分类中，我们建立了一个基于潜在空间表示的分类器。通过共享相同的潜在社会空间，两个目标强烈耦合，最后DDRW通过使用梯度下降优化连接目标。

问题定义

我们考虑网络分类问题，定义为将来自给定网络的实体分类为一个集合中的一个或多个类别， $G=(V, E, Y)$ 表示一个网络， V 是顶点集合，代表网络中的实体， $E \subseteq (V \times V)$ 是边的集合，代表实体之间的关系； $Y \subseteq \mathbb{R}^{|V| \times |Y|}$ 代表实体的标签。 Y_U 是相同图 G 中未定义的标签；分类任务的目的是从已标签的数据中学习一个模型，并且生成一个标签集合 Y_P ，成为 Y_U 的预测。

传统的机器方法是学习一个权重矩阵 H 来最小化 Y_P 和 Y_U ，而 $Y_P = \mathcal{F}(X, H)$ ， \mathcal{F} 是任何已知的固定函数。在网络分类中，网络的内部结构解决了提取实体的相邻特征。于是，核心问题就变为如何描述形式特征空间内的不规则网络。

当DeepWalk应用于网络分类任务时，首先学习了用来描述拓扑结构 G 的 X ，随后学习了一个分类器 H ，一个明显的缺点是嵌入步骤不知道目标类别标签信息，并且可能学习分类不理想的嵌入。

因此提出的DDRW通过联合学习 X 和 H ，增强了DeepWalk的效果。同时使用特定网络中拓扑和标签信息，DDRW显著提高了分类精度。

判别式深度随机游走 (DDRW)

- 嵌入目标

$$\min_{\theta} \mathcal{L}_r(\theta, \alpha), \quad (1)$$

$\theta = (\theta_1, \theta_2, \dots, \theta_{|V|})$ 表示在潜在空间中嵌入的向量

α 表示原图中的拓扑结构 \mathcal{L}_r 表示嵌入表示和原拓扑结构之间的差距

然后使用随机游走和Word2Vec方法获得目标函数

$$\mathcal{L}_r(\theta, \alpha) = - \sum_{i=1}^{\tau} \frac{1}{s} \sum_{t=1}^s \sum_{-R \leq j \leq R, j \neq 0} \log p(\omega_{i,t+j} | \omega_{i,t}). \quad (2)$$

$$p(\omega_O | \omega_I) = \frac{\exp(\theta_{\omega_O}^T \hat{\theta}_{\omega_I})}{\sum_{i=1}^{|V|} \exp(\theta_i^T \hat{\theta}_{\omega_I})}, \quad (3)$$

- 分类目标

$$\min_{\theta, \beta} \mathcal{L}_c(\theta, \beta, \mathbf{y}). \quad (4)$$

$\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{|V|})$ 表示标签, β 表示随后地分类器

我们选择被称为L2正则化和L2损失支持向量分类 (SVC) 的分类器

$$\begin{aligned} & \mathcal{L}_c(\theta, \beta, \mathbf{y}) \\ &= C \sum_{i=1}^{|V|} (\sigma(1 - \mathbf{y}_i \beta^T \theta_i))^2 + \frac{1}{2} \beta^T \beta, \end{aligned} \quad (5)$$

其中, C 为正则化参数, 如果 $x > 0$, $\sigma(x) = x$, 否则 $\sigma(x) = 0$

【1】 Rong-En Fan 2008. LIBLINEAR: A library for large linear classification.
Journal of Machine Learning Research

● 联合学习

我们方法的主要目标是对给定的网络中未标签的顶点分类。我们在中间嵌入，即潜在的表示网络结构的帮助下达到这个目标。我们同时优化第3.1节和第3.2节中的两个目标。特别的定义目标函数：

$$\mathcal{L}(\theta, \beta, \alpha, y) = \eta \mathcal{L}_r(\theta, \alpha) + \mathcal{L}_c(\theta, \beta, y).$$

其中， η 是平衡两个目标函数权重的重要参数。我们解决这个联合优化问题：

$$\min_{\theta, \beta} \mathcal{L}(\theta, \beta, \alpha, y). \quad (6)$$

$$\begin{aligned}\boldsymbol{\theta} &\leftarrow \boldsymbol{\theta} - \delta \frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}} = \boldsymbol{\theta} - \delta \left(\eta \frac{\partial \mathcal{L}_r}{\partial \boldsymbol{\theta}} + \frac{\partial \mathcal{L}_c}{\partial \boldsymbol{\theta}} \right), \\ \boldsymbol{\beta} &\leftarrow \boldsymbol{\beta} - \delta \frac{\partial \mathcal{L}}{\partial \boldsymbol{\beta}} = \boldsymbol{\beta} - \delta \frac{\partial \mathcal{L}_c}{\partial \boldsymbol{\beta}},\end{aligned}\tag{7}$$

δ 是梯度下降算法中的学习率，在我们的实现中， δ 初始设置为 0.025，随着步骤线性减少，其他参数与word2vec的默认设置相同

实验设置

● 数据集

BlogCatalog: 一个由博客作者提供的社交关系的网络，此图的标签是上传用户指定的主题

Flickr: Flickr照片共享网站用户之间的联系网络，此图的标签代表用户对某些类别照片的兴趣

YouTube: Youtube视频分享网站用户之间的网络。 标签代表对不同类型的视频感兴趣的用戶组

● 基本方法

LINE、DeepWalk

Spectral Clustering: 该方法发现图切割对分类任务很有用。【1】

Edge Cluster: 该方法使用k均值聚类算法将图形的边缘分割成几部分。然后它在小群集上运行迭代以分别查找内部关系【2】

Majority: 只是选择了使用频率最高的标签，没有使用任何图的结构信息

实验

● 分类任务地效果

	Labeled Nodes	10%	20%	30%	40%	50%	60%	70%	80%	90%
Micro- F_1 (%)	DDRW	37.13	39.31	41.08	41.76	42.64	43.17	43.80	44.11	44.79
	LINE	35.42	37.89	39.71	40.62	41.46	42.09	42.55	43.26	43.68
	DeepWalk	36.00	38.20	39.60	40.30	41.00	41.30	41.50	41.50	42.00
	SpecClust	31.06	34.95	37.27	38.93	39.97	40.99	41.66	42.42	42.62
	EdgeClust	27.94	30.76	31.85	32.99	34.12	35.00	34.63	35.99	36.29
	Majority	16.51	16.66	16.61	16.70	16.91	16.99	16.92	16.49	17.26
Macro- F_1 (%)	DDRW	21.69	24.33	26.28	27.78	28.76	29.53	30.47	31.40	32.04
	LINE	20.98	23.44	24.91	26.06	27.19	27.89	28.43	29.10	29.45
	DeepWalk	21.30	23.80	25.30	26.30	27.30	27.60	27.90	28.20	28.90
	SpecClust	19.14	23.57	25.97	27.46	28.31	29.46	30.13	31.38	31.78
	EdgeClust	16.16	19.16	20.48	22.00	23.00	23.64	23.82	24.61	24.92
	Majority	2.52	2.55	2.52	2.58	2.58	2.63	2.61	2.48	2.62

Table 2: Multi-class classification results in BlogCatalog.

实验

● 参数敏感度

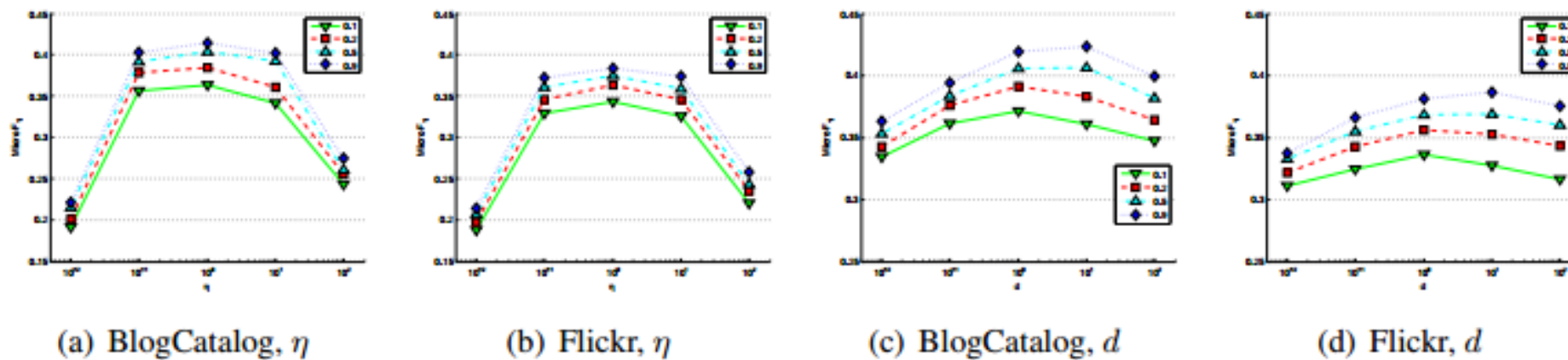


Figure 3: Parameter Sensitivity in BlogCatalog and Flickr