洛伦兹插值

In the following, we describe our approach for learning continuous hierarchies from unstructured observations.

双曲几何与Poincare插值

双曲空间是一个完备、独特的简单负曲率黎曼流形。我们知道一些等价的双曲模型，可以根据任务来选择哪个最合适。

庞加莱球：

$$\begin{aligned}\mathcal{P}^n &= (\mathcal{B}^n, g_p) \\ \mathcal{B}^n &= \{x \in \mathbb{R}^n : ||x|| < 1\}. \\ g_p(x) &= (\frac{2}{1-||x||^2})^2 g_e \\ d_p(x,y) &= arcosh(1 + 2\frac{||x-y||^2}{(1-||x||^2)(1-||y||^2)}).\end{aligned} \tag{1}$$

洛伦兹模型的黎曼优化

洛伦兹双曲空间

以下，设$x, y \in \mathbb{R}^{n+1}$，并且记洛伦兹标量积为：

$$\langle x, y \rangle_{\mathcal{L}} = -x_0 y_0 + \sum_{i=1}^{n} x_n y_n \tag{2}$$

这是一个定义于n维黎曼流形$\mathcal{L}^n = (\mathcal{H}^n, g_t)$n维双曲空间. 其中

$$\mathcal{H}^n = \{x \in \mathbb{R}^{n+1} : \langle x, x \rangle_{\mathcal{L}} = -1, x_0 > 0\} \tag{3}$$

记$g_l(x) = diag\{-1, 1, 1, 1, \cdots, 1\}$. 距离函数为

$$d_l(x, y) = arcosh(-\langle x, y\rangle_{\mathcal{L}}). \tag{4}$$

特别的，对于任意的$x = (x_0, x') \in \mathbb{R}^{n+1}$

$$x \in \mathcal{H}^n \Leftrightarrow x_0 = \sqrt{1 + ||x'||}. \tag{5}$$

黎曼优化

黎曼流形是一个有着g度量的光滑连续流形$\mathcal{M}$。对于每一个$x \in \mathcal{M}$, 记$\mathcal{T}_x\mathcal{M}$为其正切空间。度量g引入了一个内积$\langle \cdot, \cdot \rangle_x : \mathcal{T}_x\mathcal{M} \times \mathcal{T}_x\mathcal{M} \to \mathbb{R}$. 测地线$\gamma : [0, 1] \to \mathcal{M}$. 指数映射

$$\begin{gathered} \exp_x : \mathcal{T}_x\mathcal{M} \to \mathcal{M} : v \in \mathcal{T}_x\mathcal{M}, \\ \exp_x(v) = y, \gamma(0) = x, \gamma(1) = y, \\ \gamma'(0) = \frac{\partial}{\partial t}\gamma(0) = v. \end{gathered} \tag{6}$$

对于完备流形$\mathcal{M}$, 指数映射可以对任意的$x \in \mathcal{M}$定义。

我们需要考虑的问题是，寻找

$$\min_{\theta \in \mathcal{M}} f(\theta) \tag{7}$$

所以需要做的就是逐步更新θ：

$$\theta_{t+1} = \exp_{\theta_t}(-\eta \, \mathrm{grad} f(\theta_t)) \tag{8}$$

那么对于洛伦兹模型， 正切空间$\mathcal{T}_x\mathcal{L}^n$定义为所有垂直于$x$的向量的集合。也就是

$$\mathcal{T}_x\mathcal{L}^n = \{v : \langle x, v\rangle_{\mathcal{L}} = 0\}. \tag{9}$$

如果v在正切空间上，那么指数映射$\mathcal{T}_x\mathcal{L}^n \to \mathcal{L}^n$定义为

$$\exp_x(v) = \cosh(||v||_{\mathcal{L}})x + \sinh(||v||_{\mathcal{L}})\frac{v}{||v||_{\mathcal{L}}}. \tag{10}$$

为了计算公式7的参数，我们需要**黎曼梯度**。

1. 计算最大梯度的方向：

   $$h = g_l^{-1}\nabla f(\theta) \tag{11}$$

   由于g_l是对合矩阵，公式11的拟很容易求。为了从h中求得黎曼梯度，我们使用中交映射$\mathrm{proj}_\theta : \mathbb{R}^{n+1} \to \mathcal{T}_\theta\mathcal{L}^n$.

   $$\mathrm{proj}_x(u) = u - \frac{\langle x, u\rangle_{\mathcal{L}}}{\langle x, x\rangle_{\mathcal{L}}}x = u + \langle x, u\rangle_{\mathcal{L}}x \tag{12}$$

2. 利用如下的算法来估计θ, 从$\mathcal{U}(-10^{-3}, 10^{-3})$取样，用公式5来固定$x_0$.
   1. 输入学习率η, 训练次数T.
   2. 对于$t = 1, \cdots, T$,

$$
\begin{aligned}
h_t &\leftarrow g_{\theta_t}^{-1} \nabla f(\theta_t) \\
\text{grad } f(\theta_t) &\leftarrow \text{proj}_{\theta_t}(h_t) \\
\theta_{t+1} &\leftarrow \exp_{\theta_t}(-\eta \, \text{grad} f(\theta_t))
\end{aligned}
\tag{13}
$$

等价模型

3.2.3, 不表。

通过相似性推断概念层次

- $\mathcal{C} = \{c_i\}_{i=1}^m$为概念集合。
- $X \in \mathbb{R}^{m \times m}$为两两之间的相似性得分
- 假设**概念**可以由偏序关系$(\mathcal{C}, \preceq)$刻画。也就是可以排列出哪两个概念更为相似。
- 我们的目标是重建X中的偏序关系。所以以如下两种方面分解语义：
    1. 两个概念是否可比较（$c_i \sim c_j$）
    2. 是否有一个更为通用($c_j \sqsubset c_i$)

只考虑X的话，我们很难推断出$c_i \sim c_j$. 但是我们假设这是可行的。于是就可以进行局部排序。令

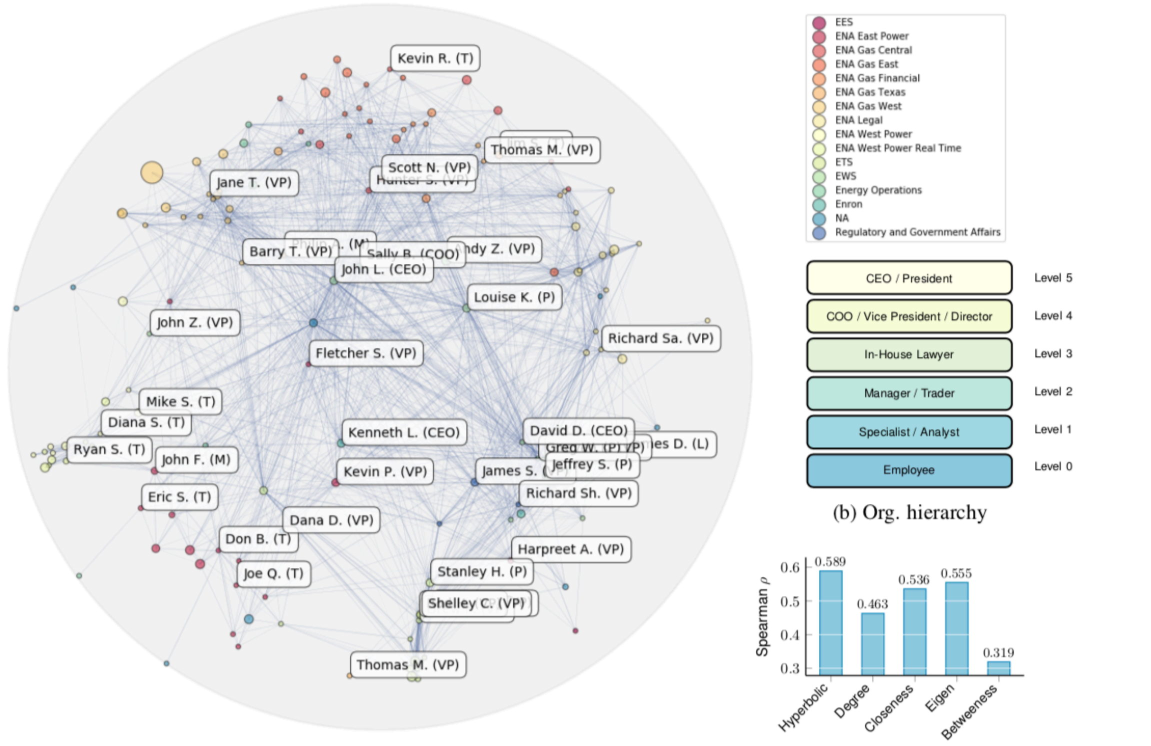
$$
\phi(i, j) = \arg_{k \in \mathcal{N}(i,j)} \min d(u_i, u_k) \tag{14}
$$

为c_i在$\mathcal{N}(i, j)$中最近的邻居，我们可以搞一个插值：$\Theta = \{u\}_{i=1}^m$, 使得

$$
\max_{\Theta} \sum_{i,j} \log P(\phi(i, j) = j | \Theta) \tag{15}
$$

这里概率$P(\phi(i, j) = j | \Theta) = \frac{e^{-d(u_i, u_j)}}{\sum_{k \in \mathcal{N}(i,j)} e^{-d(u_i, u_k)}}$.

为了方便计算，$\mathcal{N}(i, j)$是随便取的。



(a) Embedding of the Enron communication graph

(b) Org. hierarchy

(c) Rank-order correlation

Figure 2: Embedding of the Enron email corpus. Abbreviations in parentheses indicate organizational role: CEO = Chief Executive Officer, COO = Chief Operating Officer, P = President, VP = Vice President, D = Director, M = Manager, T = Trader. Blue lines indicate edges in the graph. Node size indicates node degree.

上面是公司职位网络的插值结果。

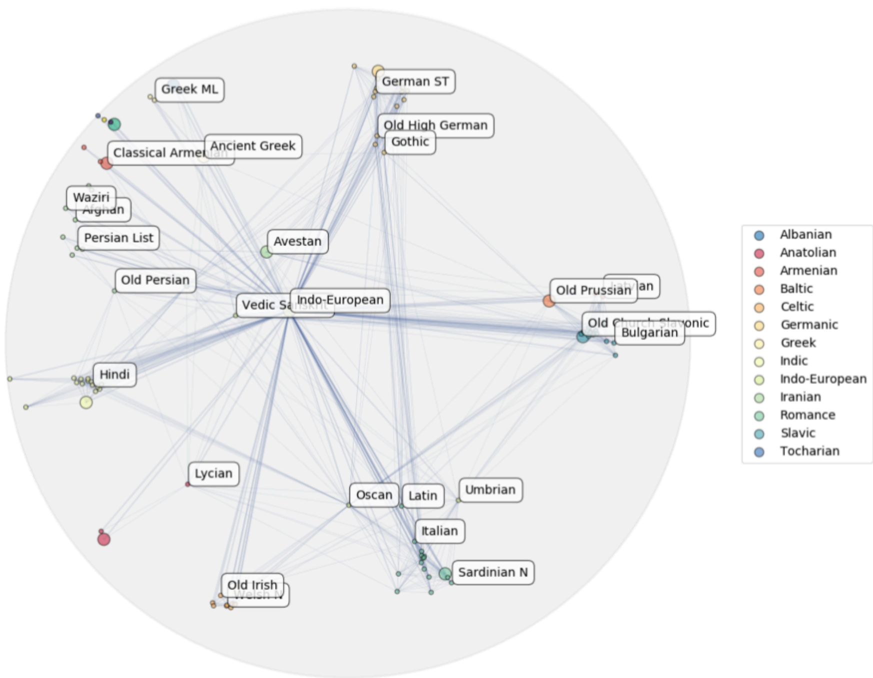


Figure 3: Embedding of the IELex lexical cognate data.

这个是语汇网络。

总结

本文更详细地讲述了很多概念的构成，和如何将插值概念与机器学习理念结合到一起。是一篇更完备的论文。我们可以看出，在一些复杂的网络结构中（如公司职位网络），一个结点可能要向多个结点汇报工作。这就会使得它的重要性得到一些提升，也同时使得相交线增多。而语义网络相对来讲，中心性很强。

wo men