

深度学习中不得不学的Graph Embedding方法

王喆

机器学习、深度学习 (Deep Learning) 话题的优秀回答者

已关注

霍华德、茶坊主、嘉慧Lincoln 等 484 人赞同了该文章

这里是「[王喆的机器学习笔记](#)」的第十四篇文章，之前已经有无数同学让我介绍一下Graph Embedding，我想主要有两个原因：

- 一是因为**Graph Embedding**是推荐系统、计算广告领域最近非常流行的做法，是从word2vec等一路发展而来的Embedding技术的最新延伸；
- 二是因为已经有很多大厂将**Graph Embedding**应用于实践后取得了非常不错的线上效果。

那我们今天就一起讨论一下Graph Embedding的主要做法和前沿应用。

word2vec和由其衍生出的item2vec是embedding技术的基础性方法（我在之前的文章中也详细介绍了二者的技术原理，[王喆：万物皆Embedding，从经典的word2vec到深度学习基本操作item2vec](#)），但二者都是建立在“**序列**”样本（比如句子、推荐列表）的基础上的。而在互联网场景下，数据对象之间更多呈现的是图结构。典型的场景是由用户行为数据生成的和物品全局关系图（图1），以及加入更多属性的物品组成的知识图谱（图2）。

▲ 赞同 484

65 条评论

分享



知乎



首发于

王喆的机器学习笔记

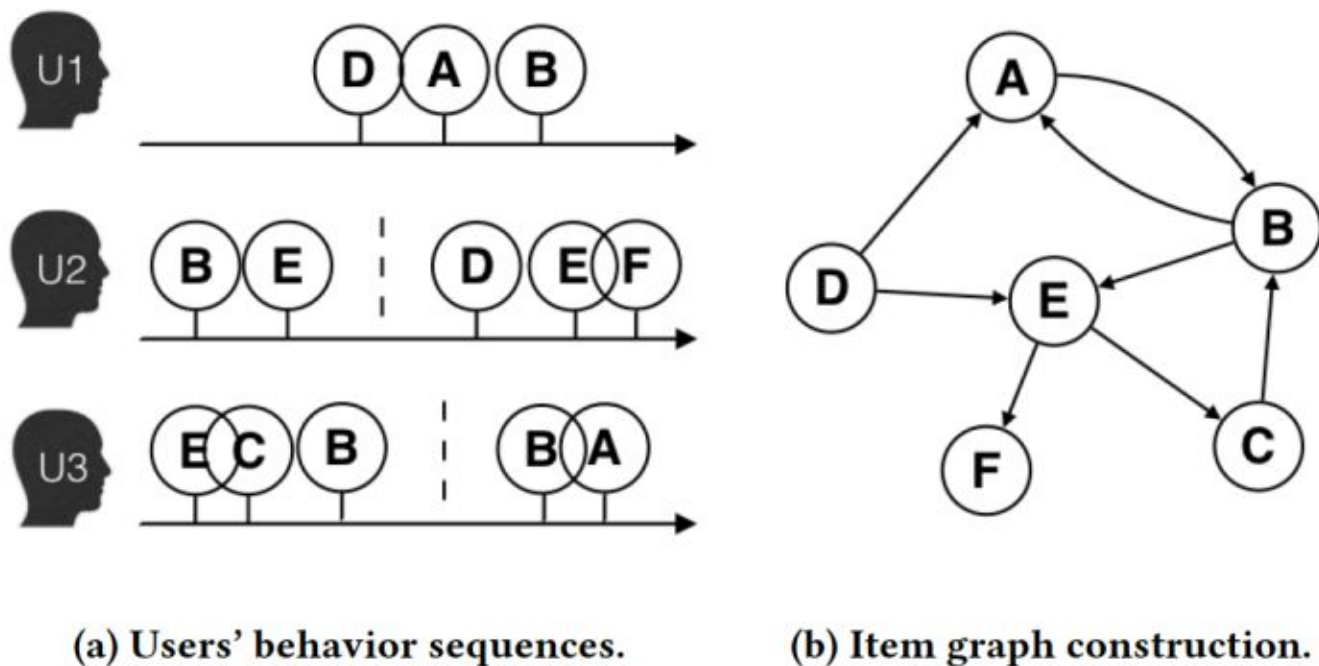


图1 由用户行为序列生成的物品全局关系图 (引自阿里论文)

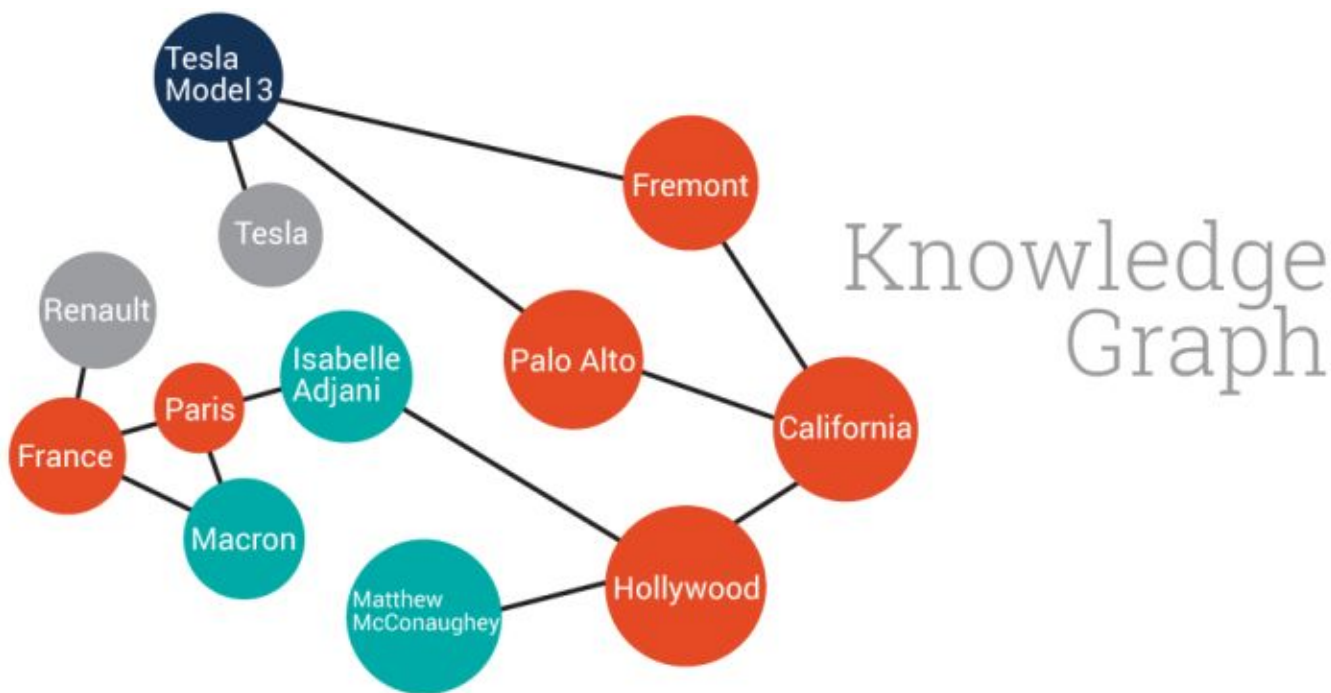


图2 由属性、实体、各类知识组成的知识图谱

在面对图结构的时候，传统的序列embedding方法就显得力不从心了。在这样的背景下，对图结构中间的节点进行表达的graph embedding成为了新的研究方向，并逐渐在深度学习推荐系统领域流行起来。

经典的Graph Embedding方法——De

▲ 赞同 484 ▼

65 条评论

➤ 分享



word2vec进行训练，得到物品的embedding。图3用图示的方法展现了DeepWalk的过程。

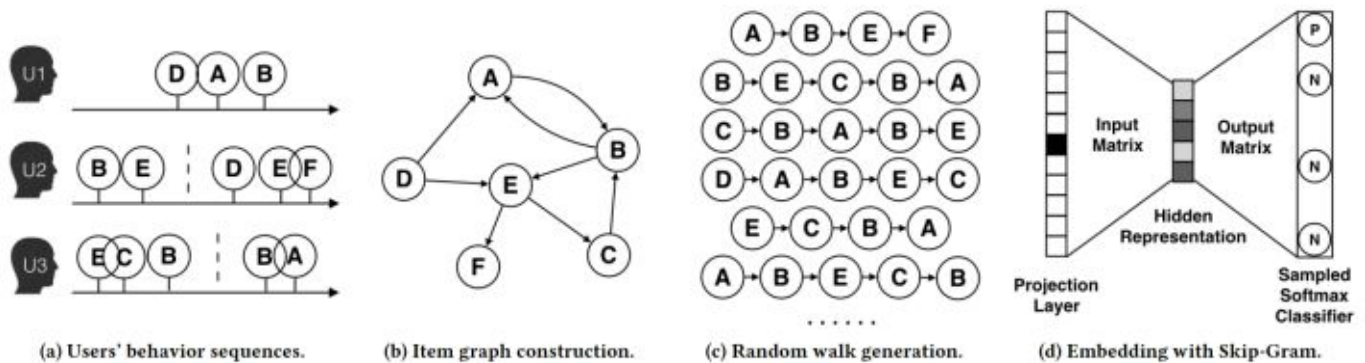


图3 DeepWalk的算法流程（引自阿里论文）

如图3，整个DeepWalk的算法流程可以分为四步：

1. 图a展示了原始的用户行为序列
2. 图b基于这些用户行为序列构建了物品相关图，可以看出，物品A，B之间的边产生的原因就是用户U1先后购买了物品A和物品B，所以产生了一条由A到B的有向边。如果后续产生了多条相同的有向边，则有向边的权重被加强。在将所有用户行为序列都转换成物品相关图中的边之后，全局的物品相关图就建立起来了。
3. 图c采用随机游走的方式随机选择起始点，重新产生物品序列。
4. 图d最终将这些物品序列输入word2vec模型，生成最终的物品Embedding向量。

在上述DeepWalk的算法流程中，核心是第三步，其中唯一需要形式化定义的是随机游走的跳转概率，也就是到达节点 v_i 后，下一步遍历 v_i 的临接点 v_j 的概率。如果物品的相关图是有向有权图，那么从节点 v_i 跳转到节点 v_j 的概率定义如下：

$$P(v_j|v_i) = \begin{cases} \frac{M_{ij}}{\sum_{j \in N_+(v_i)} M_{ij}}, & v_j \in N_+(v_i), \\ 0, & e_{ij} \notin \mathcal{E}, \end{cases}$$

其中 $N_+(v_i)$ 是节点 v_i 所有的出边集合， M_{ij} 是节点 v_i 到节点 v_j 边的权重。

如果物品相关图是无相无权重图，那么跳转概率将是上面公式的一个特例，即权重 M_{ij} 将为常数1，且 $N_+(v_i)$ 应是节点 v_i 所有“边”的集合，而不是

2016年，斯坦福大学在DeepWalk的基础上更进一步，通过调整随机游走权重的方法使graph embedding的结果在网络的**同质性（homophily）**和**结构性（structural equivalence）**中进行权衡权衡。

具体来讲，网络的“同质性”指的是距离相近节点的embedding应该尽量近似，如图4，节点u与其相连的节点s1、s2、s3、s4的embedding表达应该是接近的，这就是“同质性”的体现。

“结构性”指的是结构上相似的节点的embedding应该尽量接近，图4中节点u和节点s6都是各自局域网络的中心节点，结构上相似，其embedding的表达也应该近似，这是“结构性”的体现。

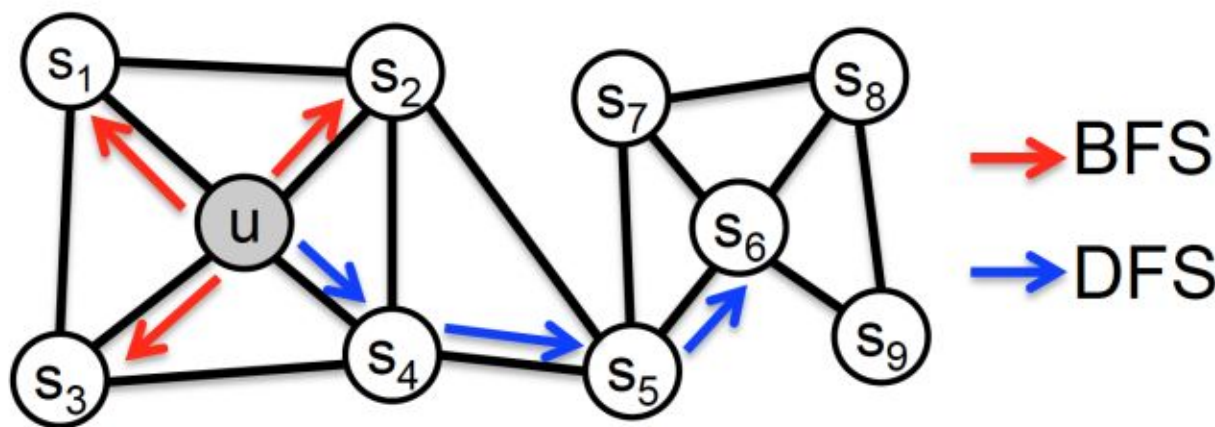


图4 宽度优先搜索（BFS）和 深度优先搜索（DFS）示意图

为了使Graph Embedding的结果能够表达网络的**同质性**，在随机游走的过程中，需要让游走的过程更倾向于**宽度优先搜索（BFS）**，因为BFS更喜欢游走到跟当前节点有直接连接的节点上，因此就会有更多同质性信息包含到生成的样本序列中，从而被embedding表达；另一方面，为了抓住网络的**结构性**，就需要随机游走更倾向于**深度优先搜索（DFS）**，因为DFS会更倾向于通过多次跳转，游走到远方的节点上，使得生成的样本序列包含更多网络的整体结构信息。（通过 @张备 同学的提醒，这里其实是写反了，BFS应该反映了结构性，DFS反而反应了同质性，大家可以深度思考一下这是为什么，欢迎在评论区讨论）

那么在node2vec算法中，是怎样控制BFS和DFS的倾向性的呢？主要是通过节点间的跳转概率。图5显示了node2vec算法从节点t跳转到节点v后，下一步从节点v跳转到周围各点的跳转概率。

知乎



首发于

王喆的机器学习笔记

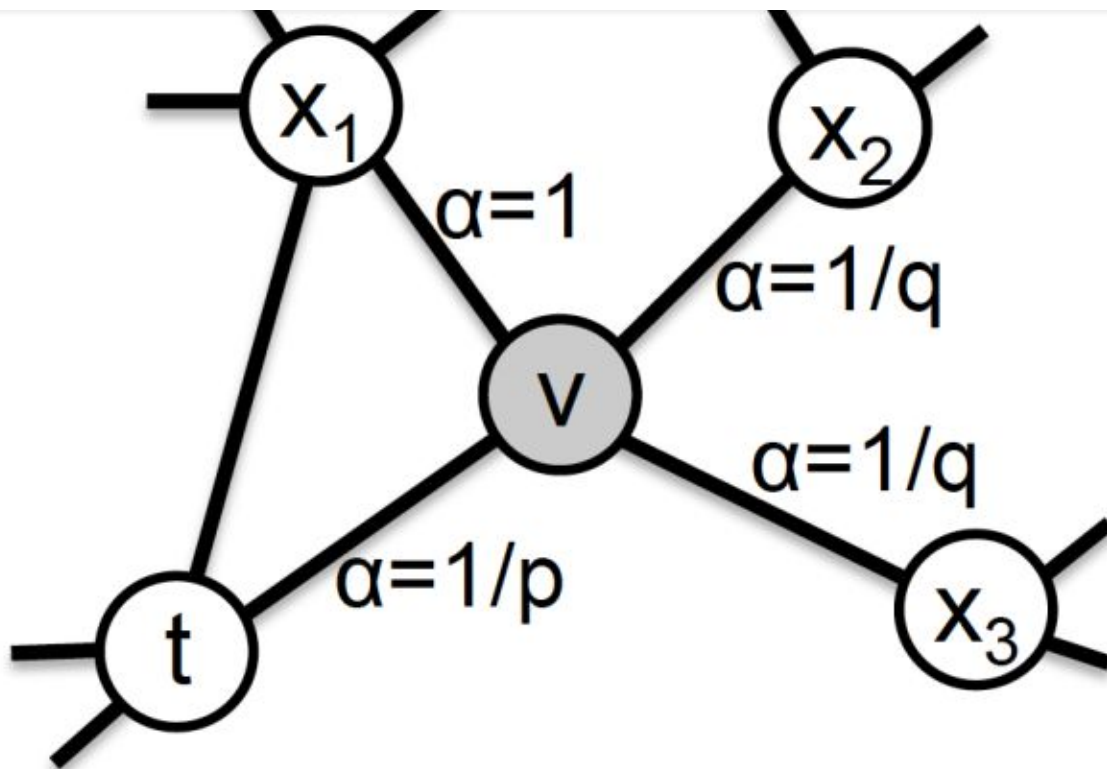


图5 node2vec的跳转概率

形式化来讲，从节点v跳转到下一个节点x的概率为

$$\pi_{vx} = \alpha_{pq}(t, x) \cdot \omega_{vx}$$

其中 ω_{vx} 是边vx的权重， $\alpha_{pq}(t, x)$ 的定义如下：

$$\alpha_{pq}(t, x) = \begin{cases} \frac{1}{p} & \text{if } d_{tx} = 0 \\ 1 & \text{if } d_{tx} = 1 \\ \frac{1}{q} & \text{if } d_{tx} = 2 \end{cases}$$

其中，dtx指的是节点t到节点x的距离，参数p和q共同控制着随机游走的倾向性。参数p被称为返回参数（return parameter），p越小，随机游

▲ 赞同 484 ▼

● 65 条评论

➤ 分享





node2vec这种灵活表达同质性和结构性的特点也得到了实验的证实。图6的上图就是node2vec更注重同质性的体现，可以看到距离相近的节点颜色更为接近，而图6下图则是结构特点相近的节点的颜色更为接近。

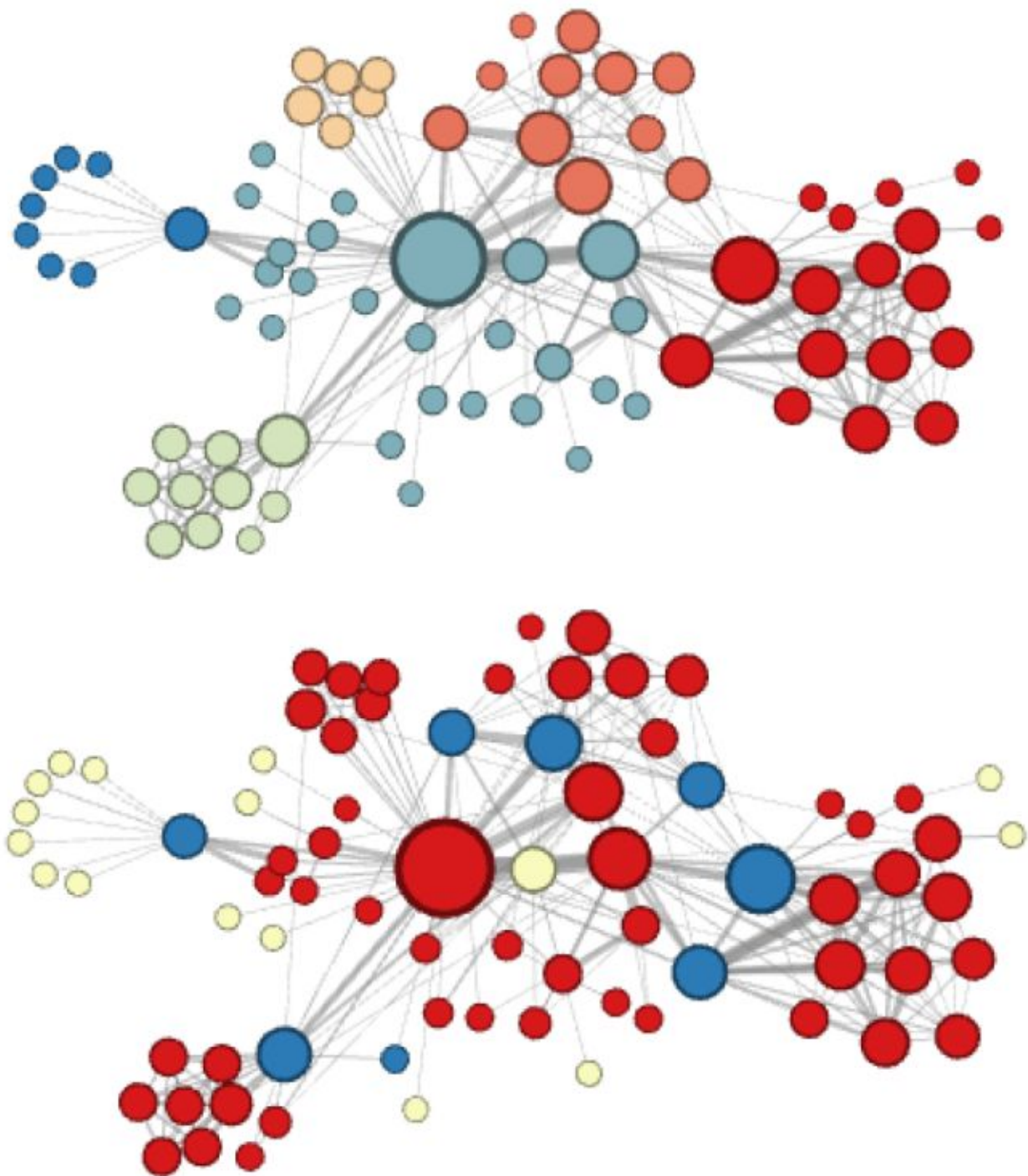


图6 node2vec实验结果

node2vec所体现的网络的同质性和结构性在推荐系统中也是可以被很直观的解释的。**同质性相同的物品很可能是同品类、同属性、或者经常被一同购买的物品，而结构性相同的物品则是各品类的爆款、各品类的最佳凑单商品等拥有类似趋势或**中都是非常重要的特征表达。由于node2vec的

▲ 赞同 484 ▼

65 条评论

分享





阿里的Graph Embedding方法EGES

2018年阿里公布了其在淘宝应用的Embedding方法EGES (Enhanced Graph Embedding with Side Information)，其基本思想是在DeepWalk生成的graph embedding基础上引入补充信息。

如果单纯使用用户行为生成的物品相关图，固然可以生成物品的embedding，但是如果遇到新加入的物品，或者没有过多互动信息的长尾物品，推荐系统将出现严重的冷启动问题。**为了使“冷启动”的商品获得“合理”的初始Embedding，阿里团队通过引入了更多补充信息来丰富Embedding信息的来源，从而使没有历史行为记录的商品获得Embedding。**

生成Graph embedding的第一步是生成物品关系图，通过用户行为序列可以生成物品相关图，利用相同属性、相同类别等信息，也可以通过这些相似性建立物品之间的边，从而生成基于内容的knowledge graph。而基于knowledge graph生成的物品向量可以被称为补充信息 (side information) embedding向量，当然，根据补充信息类别的不同，可以有多个side information embedding向量。

那么如何融合一个物品的多个embedding向量，使之形成物品最后的embedding呢？最简单的方法是在深度神经网络中加入average pooling层将不同embedding平均起来，阿里在此基础上进行了加强，对每个embedding加上了权重，如图7所示，对每类特征对应的Embedding向量，分别赋予了权重 $a_0, a_1 \dots a_n$ 。图中的Hidden Representation层就是对不同Embedding进行加权平均操作的层，得到加权平均后的Embedding向量后，再直接输入softmax层，这样通过梯度反向传播，就可以求的每个embedding的权重 $a_i (i=0 \dots n)$ 。



知乎



首发于

王喆的机器学习笔记

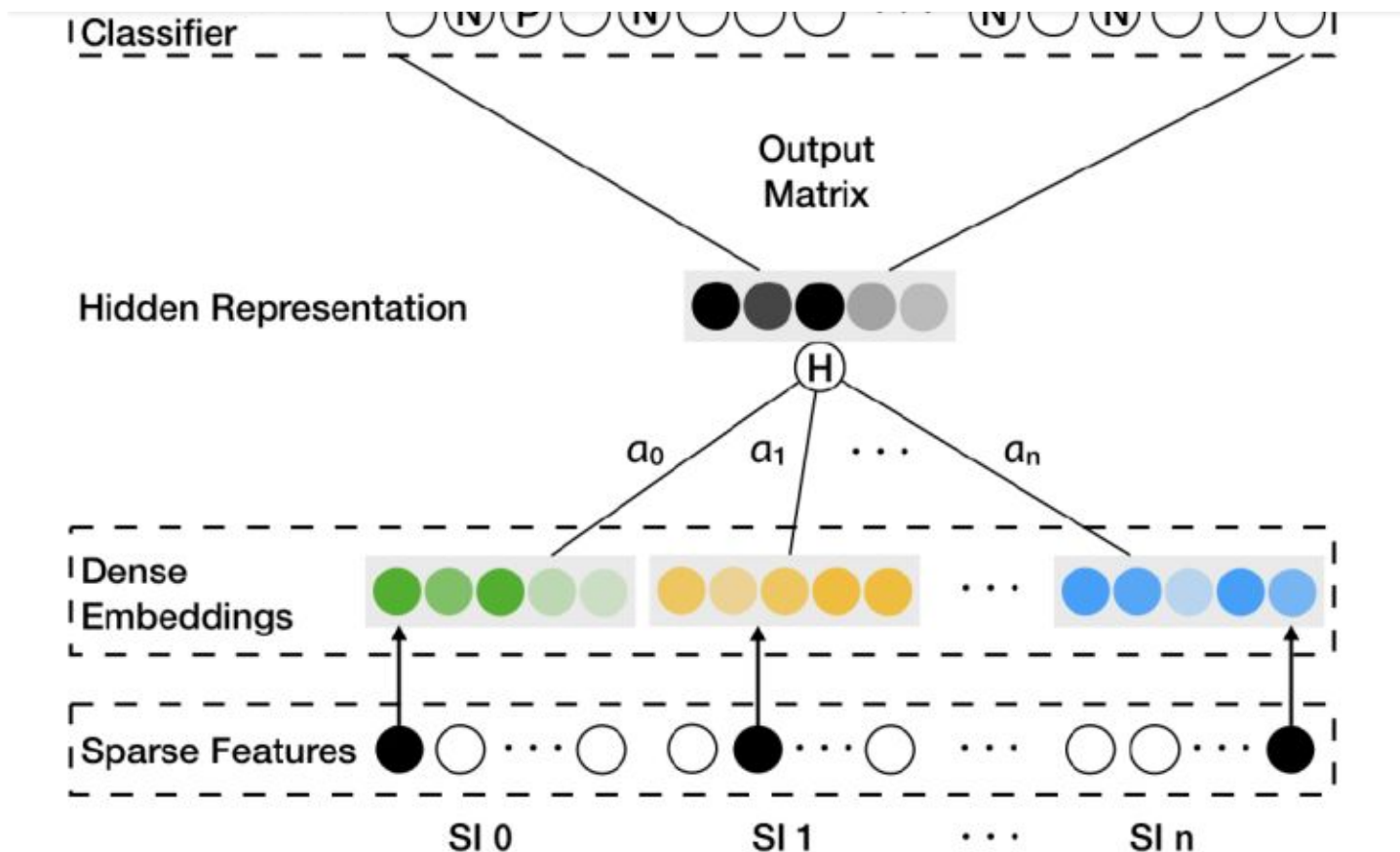


图7 阿里的EGES Graph Embedding模型

在实际的模型中，阿里采用了 e^{a_j} 而不是 a_j 作为相应embedding的权重，一是避免权重为0，二是因为 e^{a_j} 在梯度下降过程中有良好的数学性质。

阿里的EGES并没有过于复杂的理论创新，但给出一个工程性的结合多种Embedding的方法，降低了某类Embedding缺失造成的冷启动问题，是实用性极强的Embedding方法。

时至今日，Graph Embedding仍然是工程和学术领域研究和实践的热点，除了本节介绍的DeepWalk，Node2vec，EGES等主流的方法，还有LINE，SDNE等方法也很流行，感兴趣的同学可以参考我之前的综述性文章 [王喆：Embedding从入门到专家必读的十篇论文](#) 或者直接阅读论文。

按照惯例留一个讨论题，希望大家踊跃发表意见，让我们所有人都从讨论中受益：

在使用Graph Embedding生成物品的Embedding之后，应该如何生成用户的Embedding？除了Average Pooling有更好的方法吗？

最后欢迎大家关注我的微信公众号：**王喆的机器学习笔记（wangzhenotes）**，

▲ 赞同 484 ▼

65 条评论

分享



知乎



首发于

王喆的机器学习笔记



跟踪计算广告、推荐系统等机器学习领域前沿。想进一步交流的同学也可以通过公众号加我的微信一同探讨技术问题，谢谢！

参考资料：

- [王喆：万物皆Embedding，从经典的word2vec到深度学习基本操作item2vec](#)
- [王喆：Embedding从入门到专家必读的十篇论文](#)
- [\[Graph Embedding\] DeepWalk- Online Learning of Social Representations \(SBU 2014\)](#)
- [\[Node2vec\] Node2vec - Scalable Feature Learning for Networks \(Stanford 2016\)](#)
- [\[Alibaba Embedding\] Billion-scale Commodity Embedding for E-commerce Recommendation in Alibaba \(Alibaba 2018\)](#)
- [王喆的推荐系统论文列表](#)

编辑于 2019-05-06

「真诚赞赏，手留余香」

赞赏

2 人已赞赏



深度学习 (Deep Learning)

机器学习

推荐系统

▲ 赞同 484 ▼

💬 65 条评论

➦ 分享

