

神奇的Embedding

原创 banana 深度传送门 2019-10-30

作者: banana

来源: <https://zhuanlan.zhihu.com/p/53058456>

整理: 深度传送门

写在开头的话

最近被来自Google AI Language的BERT语言模型刷屏，BERT效果也的确好，我趁公司Hackathon活动，研究了一把BERT，研究BERT是否可以用到推荐系统上面，大家或许会有些疑惑，这是语言模型，怎么可以用到八杆子打不着的推荐系统上面。这篇文章并不想细致地讨论BERT的优势演变之类的话题，如果你有兴趣，来自张俊林的一篇文章[1]把BERT讲解得非常好。当前这篇文章用来总结最近对语言模型和推荐系统思考，想从一个更高层面的视野来看神奇Embedding。

神奇的Embedding

神奇的Embedding在各个领域问题当中都有应用

物品搜索：Airbnb在搜索排序当中利用Embedding[2]，此方法最早发表在KDD，题为Real-time Personalization using Embeddings for Search Ranking at Airbnb，也可以参考知乎的一篇论文解读文章[3]。

自然语言处理：我们都知道，机器学习不能直接处理人类语言，需要将其转换成向量才能处理。自然语言处理当中最基础的问题之一是如何用向量表示一句话，最经典的是Bag of Words模型，之后Google推出了Word2Vec模型，可以将语义相近的词映射到向量空间中相近的位置，比如说Girl和Lady的词向量距离会比较近，之后Google提出BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding，BERT可以考虑到相同词在不同位置会有不同含义等信息，利用这个新的语言模型刷新了问答、文本情感分析等多个语言任务的表现。更详细的语言模型的演化过程可以参见，张俊林的知乎文章[1]。

推荐：矩阵分解技术就是学习物品和用户各自的Embedding，相关的文章有Matrix Factorization Techniques for Recommender Systems[4]，这个模型获得了2006年NETFLIX大赛的冠军之后大火，另外继Word2Vec大火之后，有人将其思路应用于推荐

系统，出现了Item2Vec[5]。Word2Vec在某种程度上和Matrix Factorization的本质是相同的[6, 7]。

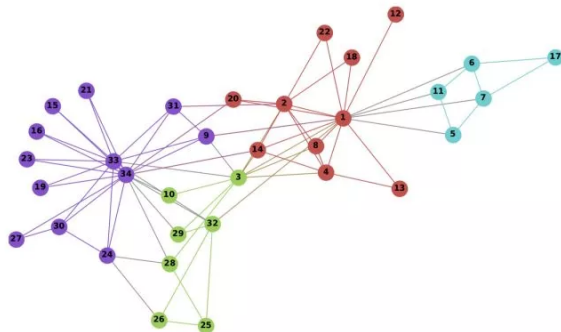
更多例子：Facebook提出Embed All The Things![8]，表示可以对任何东西都生成其Embedding，并开源了StartSpace[9]。任何可以形成网络结构的东西，都可以有Embedding的表示，Chih-Ming Chen等人[10]把相关的文章罗列出来，非常有参考价值。

形式化表示

针对不同问题，Embedding在工业实现有不同的形态，但是都可以用统一的数学符号来描述。

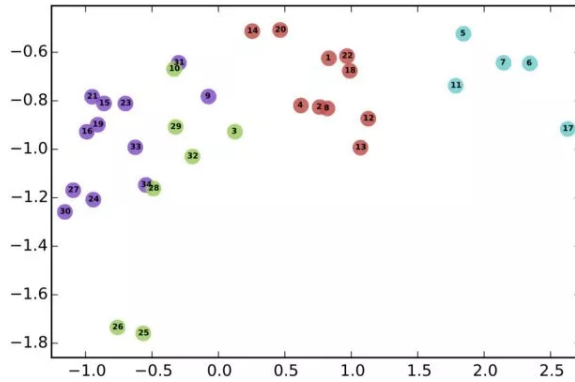
像之前的专栏文章一样，考虑机器学习当中基础的分类问题[11]，有 (X, Y) ，其中 X 是特征， Y 是类别， $X \in \mathbb{R}^{N \times S}$ ，表示有 N 条数据，每条数据有 S 个特征。 $Y \in \mathbb{R}^{N \times |\gamma|}$ ，其中 γ 为类别的集合，因为这 N 条数据之间还有额外的信息，它们互相之间有关联并构成一个网络，可以用 G 表示， $G = (V, E)$ ，其中 V 表示这 N 条数据， $|V| = N$ ， $E \subseteq (V \times V)$ ，在传统的机器学习当中可以学习到一种映射将 X 映射到类别 γ ，因为有网络 G 这样重要的信息，可以利用 G 学习 N 条数据的Embedding，也就是其在网络当中位置的表示 $X_E \in \mathbb{R}^{N \times d}$ ，对于每条数据在网络当中的位置，可以用一个 d 维的向量来表示，希望利用这 d 维向量提高最终分类任务的表现[12]。

最典型的例子是社交网络，可以通过学习对网络当中的结点的向量表示，来提升社交网络的一些机器学习任务。下面是DeepWalk[12]的例子。下图当中各个结点可以代表用户，连接关系代表是否相互为好友关系，结点上面的数字可以作为用户的ID，颜色没有特别的含义，同样的颜色都是人工标识出相互之间在空间上比较接近的一些结点。通过对这个网络的学习，得到各个结点的向量表示。



(a) Input: Karate Graph

这里的各个结点可以代表用户，连接关系是否相互为好友关系，结点上面的数字可以作为用户的ID，颜色没有特别的含义，同样的颜色都是人工标识出相互之间在空间上比较接近的一些结点



(b) Output: Representation

以二维Embedding表示结点在网络当中的位置的例子，我们希望得到，在网络当中相互接近的结点，在Embedding空间也会相互接近

如上图以二维Embedding表示结点在网络当中的位置为例，我们希望得到在Embedding空间也会相互接近的结点，在网络当中也会相互接近。

顺着这个思路，很多理论就很好解释了。

物品搜索：利用物品与用户的点击关系网络，或者利用物品与物品是否共同出现在某个用户的历史访问列表当中的物品共现关系网络，可以提取其Embedding，为物品搜索问题提供更好的线索，提升搜索结果的点击率等指标。

自然语言处理：Word2Vec利用词与词之间是否在一个句子当中出现等关系网络，提取词的Embedding表示，来获取词的更多特征，提升各类自然语言处理的表现。Bag of Words模型则只利用了词自有的特征。对于最新的BERT模型，在某种层面上相比于Word2Vec的无向网络，则考虑了词与词之间的有向网络，虽然BERT模型与网络当中的Embedding表示理论没有很显然的关系，但是我总觉得往这个方向思考的确会有一些成果。

推荐：和物品搜索问题一致，可以利用物品与物品之间的关系网络，物品和用户之间的关系网络，生成其Embedding表示，利用Embedding可以提升推荐效果。甚至可以对物品的类别，标签等信息生成Embedding[13]，其作者也开源了相关代码[14]。

结论

Embedding是一种典型的利用无监督信息提升监督问题解决效果的手段，Embedding不但在自然语言处理领域管用，在推荐系统领域也大放异彩。

如果有耐心读到这里的话，就关注一下公众号吧：)



参考资料

1. 张俊林：从Word Embedding到Bert模型—自然语言处理中的预训练技术发展史
2. Airbnb爱彼迎技术团队：Airbnb爱彼迎房源排序中的嵌入(Embedding)技术
3. 吴海波：不一样的论文解读2018 KDD best paper: Embeddings at Airbnb
4. Matrix Factorization Techniques for Recommender Systems
5. Item2Vec: Neural Item Embedding for Collaborative Filtering
6. Neural Word Embedding as Implicit Matrix Factorization
7. 吴海波：一个值得讨论的问题:word2vec与SVD/LSA等的关系
8. StarSpace: Embed All The Things!
9. facebookresearch/StarSpace
10. chihming/awesome-network-embedding
11. banana: 类别预测, Sigmoid function
12. DeepWalk: Online Learning of Social Representations
13. Metadata Embeddings for User and Item Cold-start Recommendations
14. LightFM: <https://github.com/lyst/lightfm/>



你点的每个“在看”，我都认真当成了喜欢

阅读原文