## 2018\_10\_23 对所有目前为止所看过(与中心论文有关)的论文的回顾与总结

1、A Neural Probabilistic Language Model（2003）：提出了神经语言概率模型（NNLM），用来计算词的分布式表示，避免了维度灾难，利用光滑函数的性质保证了相似词具有相似的词向量。

2、Supervised Random Walks（2011）：SRW与Pagerank算法类似，唯一的区别在于转移矩阵的构造不同。Pagerank是通过完全随机的random walk进行的，而SRW则是将数据集分为了两块：D和L，其中D中的节点更容易被s节点访问，L中的节点不容易被s节点访问（这个集合的划分是和s节点有关的），区别对待两种不同的节点，带有bias的去进行整个网络的访问，最终实现收敛pT=pT\*Q，其中Q是转移矩阵，p向量表示每个节点可能被s节点访问的概率（即是与s之间存在边的概率，用来进行边的预测）。

3、Efficient Estimation of Word Representations in Vector Space（2013）：提出了word2vec方法，CBOW模型以及Skip-gram模型，使用分层softmax进行加速训练。提出了类比推理任务。

4、Distributed Representation of Words and Phrases and their Compositionality（2013）：提出了负采样来代替分层softmax方法进行加速计算，并且针对频率较高的词以及频率较低的词进行了权衡，进一步提高了词向量的质量。（使用负采样能够使用部分样本计算来代替全局样本的条件概率的计算）

5、word2vrc Explained（2014）：详细的说明了word2vec中负采样的原理。

6、Translating Embedding for Modeling Multi-relational Data（2013）：知识图谱嵌入的基础论文**TransE**。TransE模型受到了word2vec模型结论的启发，语义具有平移性。类似的TransE模型假设知识图谱的三元组<h,r,t>的embedding向量具有平移性，即是**h**+**r**=**t**。

7、Knowledge Graph Embedding by Translating on Hyperplanes（2014）： TransE只能建模1-1的关系的原因是对于所有的关系，图中的节点的嵌入表达都是一样的。因此提出了**TransH**模型，将节点的嵌入进行投影，投影到所有计算的关系所在的平面上，这样对于每个关系而言，他们的节点的嵌入向量就不相同了，一定程度上解决了TransE的无法表示1-N，N-N，N-1关系的问题。

8、Distributed Representations of Sentences and Documents（2014.5）：将段落或者整个文件映射层一个低维密集向量。

9、**DeepWalk（2014.6）**：首先对算法的有效性进行了说明：在graph中，当它的顶点的度服从幂率分布的时候，随机游走中顶点出现的频率也服从幂率分布，而又因为自然语言处理中单词出现的频率遵循类似的幂率分布，因此该算法是可行的。之后首先对图上的每个节点进行随机游走，之后在利用NLP中词嵌入的方式，得到节点的embedding向量。（需要注意的是，这样的方法仅仅适合于**同构图**，这个方法相当于是提供了同构图嵌入的一个框架）

10、Convolutional Neural Networks for Sentence Classification（2014.9）：将词向量与CNN结合，使用CNN来提取句子的特征，之后再使用该特征来进行句子的分类预测（正/负）。

11、Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling（2014.10）：对于普通的RNN因为有梯度消失或者爆炸的问题，很难抓住长期依赖，因此出现了LSTM与GRU。该文章对GRU以及LSTM进行了详细的说明。LSTM有3个门（输入门、输出门、遗忘门）以及一个记忆单元。GRU中只有两个门（重置门、更新门）。

12、**LINE（2015.3）**：定义了一阶近似（在网络中两个节点之间的一阶近似就是他们之间的权值）。定义了二阶近似（将每个节点与网络中的其他节点之间的权值设置成一个向量p，二阶近似即是两个节点之间p向量之间的相似程度）。一阶近似反映的是网络的局部特征，二阶近似反映的是网络的全局特征。分别定义一阶相似与二阶相似的目标函数，再分别优化这两个函数即可，最后将优化得到的一阶embedding向量与二阶embedding向量连接起来。该方法仅仅适用于**同构图**。

13、**PTE Predictive Text Embedding through Large-scale（2015.8）**:利用了带标签的数据，针对与某一特定的任务训练出词向量。相比于无监督的数据训练出来的词向量而言，该方法在所训练的特定任务上效果更好，但是缺陷是训练好的embedding向量不适合除了该特定任务的其他任务。在某些特殊的数据集上，性能甚至还要超过CNN等复杂网络。定义了三个网络word-word，word-document，word-label，其中前两个都是无标签的词嵌入，而最后一个网络是利用了数据集中的标签信息。它最终是利用词向量来得到的text向量，但是它只是简单的text中所有词向量的算数平均值而已。**PTE是对LINE算法的扩展，LINE算法只适用于同构图，而PTE算法适用于异构图。（异构图中的节点是word，document）**

14、Learning Entity and Relation Embeddings for Knowledge Graph Completion（2015）：一个节点可能有多个属性，不同的关系所关注的属性不同，这就使得仅仅在同一空间内对它们进行建模是不够的。因此将节点所处的空间与关系所处的空间给分开。首先将实体投影到关系所在的空间，再使他们满足平移性。

15、Heterogeneous Network Embedding via Deep（2015）：该文章所说的异构图是指的节点为图像、文本所构成的异构图。使用CNN提取图像特征使用全连接的DNN来提取文本特征，将他们的特征映射到同一个空间内，再利用网络的特征“相近的节点的相似度越高”，构造关于特征的损失函数。

16、DAG-Structured Long Short-Term Memory for Semantic Compositionality（2016）：使用DAG结构的LSTM来解决惯用短语的非组合性问题。

17、**node2vec（2016）**：介绍了传统的BFS和DFS的搜索策略的缺点，利用**带偏好的随机游走**（兼顾了BFS与DFS的优点）在图中进行采样，利用类似与Skip-gram模型的方法表示“具有相似邻居的节点的embedding相似”进行嵌入向量的学习。（带偏好的random walk+Skip-gram）

18、Semantic proximity search on graphs with metagraph-based learning（2016）：