

# 知识表示学习演讲稿

今天我讲的主题是 知识表示学习

首先介绍一下 表示学习的定义。这是一篇文献中对表示学习的定义。

机器学习算法的性能依赖于原始数据的表示。学习一个好的数据表示，可以提高算法的性能。

比如文本分类算法，以前是用TF-IDF权重来表示词，只表现了词对文本的重要性，完全没有包含词的语义

现在使用表示学习学习词的向量表示，有效减少了文本分类算法的误分率。

词表示学习是一种典型的表示学习。

薛莹莹师姐和龚志远师兄都讨论过词的表示学习。我这里帮大家回顾一下。

根据词的语义是由其上下文决定的，这个语言学先验知识，通过深度学习，将词表示成连续稠密的实值向量。

这就是词的表示学习。

词的表示学习也是一种分布式表示。

孤立得看向量中的每一维，都没有明确对应的含义，而综合各维形成一个向量，则能够表示词的语义信息。

关于词表示，其中有两种较为常见的模型

**一种是连续词袋模型（CBOW），根据上下文，推测中心词的概率。**

**一种是skip-gram 模型，根据中心词，推测上下文的概率。**

**softmax分类器用来归一化输出单元的概率**

通过将词表示为向量的形式，人们发现了有趣的平移不变现象。

如图，存在相同语义关系的词对，他们在向量空间中的平移向量几乎相同。

China 和 Beijing, Japan 和 Tokyo 之间都存在着首都关系，他们之间的平移几乎相同。

这里介绍一些 知识学习的背景知识。首先介绍一下知识库。

知识库，可以看做包含E, R, S的集合。其中 E代表实体集 R代表关系集 S代表定义在实体集 和 关系集 上的 三元组集合 每个三元组 都是由 头实体h, 关系r, 尾实体t组成的。

将知识库用图的形式表示，就是我们经常听到的知识图谱。

通过利用左边的知识，我们可以构建右边的知识图谱。

其中节点 代表 实体，边 代表 实体间的语义关系

比如 李达康 是一个 头实体 王大路 是一个 尾实体 他们之间存在 资助关系。

知识表示学习，就是学习知识库中实体和关系的表示，将它们表示为连续稠密的低维向量。

有关学者，受到词向量中平移不变现象的启发，提出了翻译模型，认为在向量空间中，关系向量可以看作是头向量到尾向量的平移。

这是翻译模型的损失函数，就是计算头实体向量和关系向量与尾实体向量之间的L1或者L2距离。

$J$ 是模型需要优化的目标函数。

其中 $S'$ 是错误三元组的集合，错误三元组是通过随机替换正确三元组的头实体或者尾实体形成的。

$r$ 是错误三元组得分与正确三元组得分之间的距离。

这是原论文中TransE算法实现的步骤。

TransE算法步骤，其实也很简单。

首先对所有的实体和关系向量进行随机值初始化。

然后进入迭代循环过程。

每次迭代

选取一个大小为 $b$ 的三元组集合。

然后为每个正确的三元组，sample一个错误的三元组，形成我们需要的训练数据。

利用形成的小批量训练数据，对参数进行更新。

其中是利用批量梯度下降对参数进行更新的。

尾部是python实现TransE算法的代码。

应用：

链接预测：构建大规模知识图谱，需要不断补充实体间的关系，利用翻译模型，可以预测两个实体的关系，进行知识图谱的补全

相似度计算：利用实体的分布式表示，可以快速计算实体间的语义相似度。对于自然语言处理和信息检索的很多任务有重要意义。

通过原论文中的实验结果，我们可以明显发现。

对于 $N$  对  $1$  的关系，预测头部准确率极低

对于 $1$  对  $N$  的关系，预测尾部准确率极低

$N$ - $1$ 类型关系指的是，该类型关系中的一个尾实体平均对应多个头实体。

$1$ - $N$ 类型关系指的是，该类型关系中的一个头实体平均对于多个尾实体。

存在这种问题是由于TransE模型过于简单，仅仅是对层次关系进行建模，没有考虑到知识库中其他复杂关系。

在TransE模型中，实体在不同的关系下的表示相同。这是有缺陷的。

比如奥巴马在总统这个关系下和布什，含义相同。

但是在大部分别的关系下，奥巴马 布什 这两个实体含义 是不同的。

词的表示也存在相同的问题，词在不同的语境中，可能存在不同的语义。

TransH模型是TransE模型改进，认为一个实体在不同关系下有不同的表示。

头实体向量 $l_h$ 和尾实体向量 $l_t$ 沿法线 $w_r$ 投影到关系 $r$ 对应的超平面上。

分别用 $l_{hr}$ 和 $l_{tr}$ 表示。