

Convolutional 2D Knowledge Graph Embeddings

jtwang 2022/02/26 ¹

Dettmers T, Minervini P, Stenetorp P, et al. Convolutional 2d knowledge graph embeddings[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2018, 32(1).

提出了一个基于2D卷积操作+多层非线性变换的KG嵌入模型ConvE。它参数数量少，更鲁棒，更 expressive，在许多数据集上取得了SOTA效果，并且可以scale to large KGs。此外，还调研了WN18和FB15k数据集的test set leakage问题，新建数据集WN18RR。

Introduction

- 由于知识图谱包含数百万事实，链接预测模型的参数数量与计算成本要尽可能小，以便应用于现实场景中（link predictors should scale in a manageable way with respect to both the number of parameters and computational costs to be applicable in real-world scenarios）
- 现有的模型都利用一些简单的操作，如内积、矩阵乘法，并且限制参数数量（如DistMult使用对角阵表示关系）。这些模型simple and fast但shallow，学到的特征less expressive。要想增加特征数量（即模型的expressiveness），只能通过增加嵌入向量的维度，但是这样就不能scale to larger KG了。
- 为了增加特征数量independently of the embedding size就需要使用**multiple layers of features**。但之前一个多层结构的模型（HolE）are prone to overfit
- 所以，要解决这两个问题：
 - scaling problem of shallow architectures
 - overfitting problem of fully connected deep architectures

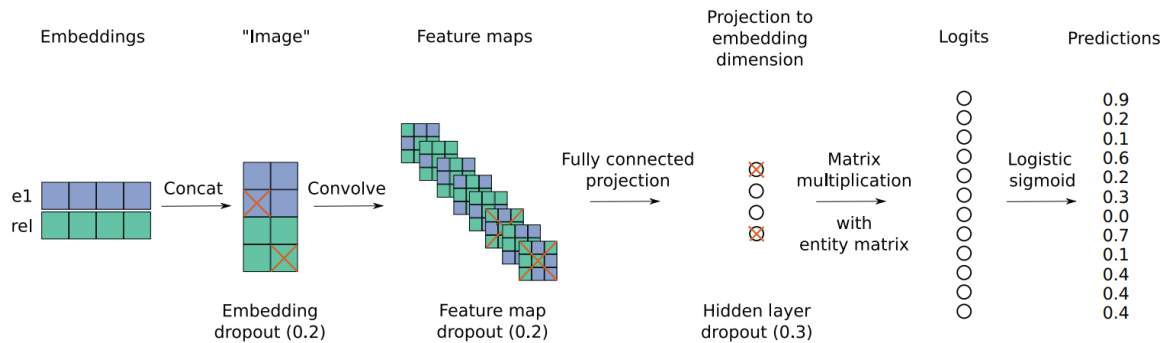
就要使用parameter efficient、fast operators which can be composed into deep networks——卷积【过渡的好自然哦，逻辑很清晰】

- 卷积操作有以下特征
 - parameter efficient, fast to compute due to GPU
 - 有许多robust methodologies控制过拟合
- 本文提出**ConvE**，使用2D卷积，是最简单的多层卷积结构：一个卷积层+映射层+内积层
- 本文的贡献
 - 介绍了一个简单、competitive的二维卷积链接预测模型ConvE
 - 开发了一个1-N score procedure，加快训练与评估速度300倍
 - 证明ConvE是highly parameter efficient，在FB15k-237数据集上获得了比DistMult和R-GCN更高的得分，但参数数量少了8x and 17x
 - the difference in performance between ConvE and a shallow model increased proportionally to the complexity of the graph【在复杂知识图谱上效果更优】
 - 调研了inverse relations test set leakage问题，提出了robust versions of datasets
 - 在robust datasets上评估ConvE和其他模型，ConvE取得了SOTA MRR

Background

- 链接预测任务 formalized as: 学习一个评分函数 ψ . *TODO*。给定一个三元组 $x = (s, r, o)$, 它的得分 $\psi(x) \in \mathbb{R}$ is proportional to the likelihood that the fact encoded by x is true
- neural link predictors 可以看做一个多层神经网络, 包含两部分
 - encoding component 【编码】: map entities \rightarrow distributed embedding representations
 - scoring component 【评分】: 一个评分函数 $\psi(s, r, o) = \psi_r(e_s, e_o) \in \mathbb{R}$

Constitutional 2D Knowledge Graphs Embeddings



- 评分函数

$$\psi_r(\mathbf{e}_s, \mathbf{e}_o) = f(\text{vec}(f([\bar{\mathbf{e}}_s; \bar{\mathbf{r}}_r] * \omega)) \mathbf{W}) \mathbf{e}_o$$

- $\bar{\mathbf{r}}_r \in \mathbb{R}^k$: a relation parameter depending on r
 - $\bar{\mathbf{e}}_s, \bar{\mathbf{r}}_r$: 2D reshaping of $\mathbf{e}_s, \mathbf{r}_r$
 - ω : filter
- 损失函数 logistic sigmoid function

$$p = \sigma(\psi_r(\mathbf{e}_s, \mathbf{e}_o))$$

$$\mathcal{L}(p, t) = -\frac{1}{N} \sum_i (t_i \cdot \log(p_i) + (1 - t_i) \cdot \log(1 - p_i))$$

- t : the label vector. $t=1$ for relationships that exists, otherwise $t=0$
- 一些细节
 - 使用非线性 f for faster training;
 - batch normalization after each layer to stabilize/regularize/increase 收敛率;
 - dropout
 - Adam optimizer
 - label smoothing 【这个是什么? ? 】
- 1-N scoring
 - 在ConvE中, 卷积操作占用75%~90%的计算时间
 - 之前的模型使用1-1 scoring: take an entity pair (s, o) and a relation as a triple, and score it
 - 1-N scoring: take one (s, r) pair and score it against all entities $o \in \mathcal{E}$
 - 1-N scoring 适合卷积网络前向-反向传播, 加快了训练和评估的速度

Experiments

- 数据集
 - WN18: 18个关系、40943个实体、151442个三元组, hierarchical structure
 - FB15k: 1345个关系、14951个实体, 三元组大多是关于电影、演员、奖项、体育
 - YAGO3-10: 37个关系、123182个实体, 三元组大多是关于人的属性, 如国籍、性别、职业
 - Countries: 用来评估模型的long-range dependencies between entities and relations, 包含3个难度递增的子任务, 最小的path length从2增加到4
 - WN18和FB15k suffer from test leakage: 测试集中的许多三元组可以简单地通过inverting训练集中的三元组得到。例如, 测试集中有 $(s, hyponym, o)$, 训练集中却包括 $(o, hypernym, s)$ 。所以, 提出了FB15k-237, 移除了inverse relations。
 - 我们提出了一个简单的rule-based模型, 在WN18和FB15k上取得了SOTA结果, 证明了test leakage这个问题的严重性
 - 创建了新的数据集WN18RR, 包含11个关系、40943个关系、93003个三元组
- Inverse model
 - 一个简单的基于规则的模型, 只建模inverse relations
 - 从训练集中抽取inverse relationships: 给定关系对 $r_1, r_2 \in \mathcal{R}$, 判断 (s, r_1, o) 是否imply (o, r_2, s)

Results

- parameter efficiency: ConvE is 17x parameter efficient than R-GCNs and 8x than DistMult
- 关于Indegree和PageRank的分析
 - Indegree: ConvE可以建模更复杂的图谱 (FB15k和FB15k-237), shallow models 如DistMult可以建模less complex KG (WN18和WN18RR)
 - PageRank: high connectivity graphs, ConvE表现更好
 - 总之, 越复杂的模型ConvE的优势越能凸显

1. 【】 里是我的注释 [🔗](#)