

JCapsR: 一种联合胶囊神经网络的藏语知识图谱表示学习模型

孙媛^{1,2,*} 梁家亚^{1,2} 陈安东^{1,2} 赵小兵^{1,2}

¹中央民族大学 信息工程学院, 北京 100081

²国家语言资源监测与研究少数民族语言中心

*Corresponding author: Yuan Sun

tracy.yuan.sun@gmail.com, lly_muc@qq.com, 903429972@qq.com, nmzxb_cn@163.com

摘要

知识图谱表示学习是自然语言处理的一项关键技术, 现有的知识图谱表示研究主要集中在英语、汉语等语言, 而低资源语言的知识图谱表示学习研究还处于探索阶段, 例如藏语。本文基于前期构建的藏语知识图谱, 提出了一种联合胶囊神经网络(JCapsR)的藏语知识图谱表示学习模型。首先, 我们使用TransR模型生成藏语知识图谱的结构化信息表示。其次, 采用融合多头注意力和关系注意力的Transfomer模型表示藏语实体的文本描述信息。最后, 采用JCapsR进一步提取三元组在知识图谱语义空间中的关系, 将实体文本描述信息和结构化信息融合, 得到藏语知识图谱的表示。实验结果表明, 相比基线系统, 联合胶囊神经网络JCapsR模型提高了藏语知识图谱表示学习的效果, 相关研究为其它低资源语言知识图谱表示学习的拓展优化提供了参考借鉴意义。

关键词: 藏语知识图谱; 表示学习; 胶囊神经网络; Transformer模型

JCapsR: A Joint Capsule Neural Network for Tibetan Knowledge Graph Representation Learning

Yuan Sun^{1,2,*} Jiaya Liang^{1,2} Andong Chen^{1,2} Xiaobing Zhao^{1,2}

¹ Minzu University of China, Beijing 100081

² National Language Resource Monitoring & Research Center of Minority Language

*Corresponding author: Yuan Sun

tracy.yuan.sun@gmail.com, lly_muc@qq.com, 903429972@qq.com, nmzxb_cn@163.com

Abstract

The knowledge graph representation learning is a key technology in the field of natural language processing. The existing research on knowledge graph representation mainly focuses on English and Chinese, etc., while research on knowledge graph representation learning for low-resource languages is still in the exploratory stage, such as Tibetan. This paper proposes a Joint Capsule Neural Network (JCapsR) model for Tibetan knowledge graph representation based on the constructed Tibetan knowledge graph. Firstly, we use the TransR model to generate a structured information representation of the Tibetan knowledge graph. Secondly, the entity text description information representation is trained by a Transfomer model incorporating multiple-head attention and relational attention. Finally, the entity text description information representation and structured information representation are fused with the JCapsR model to obtain the final representation of Tibetan knowledge graph. The experimental results show that the JCapsR model is more effective than the baselines in Tibetan knowledge graph representation learning, which provides a reference for expanding and optimizing of knowledge graph representation learning in other low-resource languages.

Keywords: Knowledge graph of Tibetan , Representation learning , Capsule neural network , Transformer model

1 引言

伴随着互联网的蓬勃发展以及信息技术的大规模普及,知识产生的速度和积累数量已经达到前所未有的高度,如何对这些知识进行获取、学习和应用是一个值得思考的问题,知识图谱技术应运而生。知识图谱表示学习,其核心是对知识图谱的实体与关系进行表示,通过表示向量准确的获得实体和关系之间存在的联系。

知识表示学习旨在将实体与关系表示为低维连续空间中的向量,基于这一思想,研究者提出了结构向量模型如TransE (Bordes et al., 2013)、TransR (Lin et al., 2015)、TransH (Wang et al., 2014)等方法进行知识表示学习的尝试,并在汉语和英语知识图谱上取得了不错的效果。而对于低资源语言(如藏语),由于缺乏大规模公开的藏语知识图谱,相关研究还处于起步阶段。此外,与中、英文数据集相比,藏语知识图谱数据比较稀疏。例如,在英文数据集FB15K (Bordes et al., 2013)中,有98%的实体具有两个或多个三元组,一个实体平均拥有39个三元组。而在藏语数据集TD50K (Sun et al., 2021)中,仅有48%的实体包含两个或多个三元组,一个实体平均仅有2个三元组。另一方面,藏语中同样存在着语义歧义。例如,藏语“ལྔ་ལྔ”在下面三个句子中分别含有“头”、“首领”和“顶端”三种不同的语义,句子1:“ཁོའི་ལྔ་ལྔ་ནས་ལྔ་ལྔ་བྱུང་།”(他的头上流血了);句子2:“མཐོང་ནི་མཐོང་ལྔ་ལྔ་”(才让是部落的首领);句子3:“ལྔ་ལྔ་ལྔ་ལྔ་ལྔ་ལྔ་”(舌头的顶端称为舌尖)。这些问题为藏语知识图谱的表示带来了更多的困难。

针对以上问题,本文提出了一种联合胶囊神经网络JCapsR (Joint Capsule Neural Network)模型对藏语知识图谱进行表示学习。首先,我们使用TransR模型生成藏语知识图谱的结构化信息表示,TransR模型能够为每个关系构造相应的向量空间,能有效解决TransE模型中存在的无法区分两个语义相近的实体在某些特定关系上不同的问题。其次,采用融合多头注意力和关系注意力的Transfomer模型表示藏语实体的文本描述信息。最后,采用JCapsR进一步提取三元组在知识图谱语义空间中的关系,将实体文本描述信息和结构化信息融合,得到藏语知识图谱的表示。该模型提高了藏语知识图谱表示学习的效果。本文的主要贡献如下:

(1) 针对藏语的语义歧义问题,本文采用TransR模型对藏语的三元组结构化信息进行表示,能够更细致的将不同的关系作为不同的投影空间,更加充分表示实体与关系的语义联系。

(2) 为了更好的表示实体的描述信息,采用融合多头注意力和关系注意力的Transfomer模型,可以更好的将文本描述信息与实体关系相关联,让实体描述信息关注到三元组关系特征,捕捉到更加多元的特征,一定程度解决了藏语知识图谱数据稀疏的问题。

(3) 为了更好的表示知识图谱中的三元组,我们提出了一种联合胶囊神经网络JCapsR模型,将联合模型和胶囊神经网络融合。其中,联合模型采用平移假设思想合并实体文本描述信息和三元组结构化信息。然后,胶囊神经网络对融合后的三元组向量进行优化,利用其优越的空间信息提取能力,提取更多的语义特征,更深度的挖掘三元组在知识图谱语义空间中的关系。

2 相关工作

2012年,谷歌提出了知识图谱的概念,随后出现了一批大规模公开的知识库,如WikiData (Vrandečić and Krotzsch, 2014)、DBpedia (Auer et al., 2007)和Freebase (Bollacker et al., 2008)等。知识表示学习旨在将实体和关系表示为稠密低维实值向量,从而在低维空间中高效计算实体、关系及其之间的复杂语义关联,对知识库的构建、推理以及融合均具有重要的意义。在知识表示学习模型方面,2013年Bordes等人提出了翻译模型TransE表示知识库,将知识库中的关系看作实体间的某种平移向量 (Bordes et al., 2013)。随后,针对TransE对于“1-to-N”、“N-to-1”和“N-to-N”等关系类型存在的不足,许多研究者提出了一系列的改进方法,核心思想是根据语义关系类型,设定映射函数,将三元组的头实体和尾实体映射到目标关系相关的语义空间中,再利用TransE的加性模型学习知识库的表示 (刘知远 et al., 2016) (Wang et al., 2014) (Ji et al., 2015) (Xiao et al., 2015)。2015年, Lin等人提出了融合实体描述信息知识表示学习模型 (Description-embodied Knowledge Representation Learning, DKRL) (Lin et al.,

2015), 通过卷积神经网络模型从实体描述中获取与实体相关的信息, 然后采用TransE平移假设思想将实体描述信息与三元组结构化信息进行融合, 实现了文本空间与知识空间的对齐与联合训练。

在文本表示方面, 常用的有CNN、RNN、LSTM网络。2017年, Vaswani等人提出了Transformer模型 (Vaswani et al., 2017), 突破了RNN中不能并行计算的限制, 其次相比于CNN, Transformer在计算两个词向量之间的关联所需要的操作数不会随着距离的增加而增长, 关键优势是内部的self-attention可以使得模型更具有可解释性, 多头注意力可以执行不同的任务。2017年, 为了解决CNN池化操作会主动丢弃大量信息可能导致有价值的空间信息严重丢失的问题, Hinton等人 (Hinton et al., 2018)提出了基于动态路由规则的胶囊神经网络模型, 可以利用向量的方向来表示目标的存在, 目标的特征可以由向量的长度来表示。胶囊神经网络在自然语言处理领域也有广泛的应用, Cheng等人 (Cheng et al., 2020)运用融合实体描述的多尺度胶囊嵌入模型来对知识图谱进行表示学习, 将实体的结构化信息和描述信息进行融合, 从而获得知识图谱词向量的最终表示形式, 此模型没有考虑对实体描述的关系进行词嵌入处理, 因此最终的实验效果还有待提升。Wang等人 (Wang et al., 2020)提出了融入多尺度特征注意力的胶囊神经网络在文本分类中的应用, 在文本分类任务上取得较好的效果, 说明对比传统的神经网络算法, 胶囊神经网络具有更加强大的文本特征捕捉能力。

对于藏语, 由于缺乏大规模公开的藏语知识库, 目前对藏语知识图谱的表示研究较少。在藏文知识库构建方面, 主要研究集中在领域知识库 (冯小兰 et al., 2019)以及实体关系抽取方面 (朱臻 et al., 2015) (夏天赐 et al., 2018) (王丽客 et al., 2020)。在藏语知识图谱表示方面, Sun等人提出了一种基于百科的藏语知识图谱表示学习的联合模型 (Sun et al., 2021), 采用其他语言知识图谱和POI地图信息的三元组来扩展藏语知识图谱, 创建了包含12,584个实体, 8,0178个三元组, 4,256对关系的语料, 同时利用联合模型得到了藏语的知识图谱表示。本文在以上工作的基础上, 提出了一种联合胶囊神经网络JCapseR模型, 能够提取三元组的结构化信息, 表示更丰富的抽象特征, 从而进一步提取语义特征和挖掘三元组在知识图谱语义空间中的关系, 从而能够更好的对藏语知识图谱表示进行学习。

3 整体架构

本文在前期构建的藏语知识库的基础上, 提出了一种联合胶囊神经网络JCapsR模型, 整体框架如图1所示。该模型主要由三个部分组成。

(1) 结构化信息表示: 利用TransR模型对藏语知识图谱三元组信息进行结构化表示。

(2) 实体描述信息表示: 采用Transformer模型和Attention进行实体文本描述信息表示, 并融合了关系注意力机制, 提取实体描述中包含的重要语义特征。

(3) 联合胶囊神经网络JCapsR模型: 运用平移假设思想, 对实体文本描述信息表示学习和结构化信息表示学习进行了联合训练, 然后经过胶囊神经网络的卷积操作和动态路由以及激活函数的处理, 获得藏语知识图谱的表示。

4 模型描述

4.1 结构化信息表示

本文采用了TransR模型学习藏语三元组的表示。TransR为每种关系 r 定义了单独的语义空间 \mathbb{R}^m , m 表示关系向量的维度, 给定一个三元组 (h, r, t) , 使用映射矩阵 $M_r \in \mathbb{R}^{n \times m}$, n 表示头实体 h 和尾实体 t 的维度。将实体映射到 r 所在的语义空间中, 可以得到头实体和尾实体的结构化向量表示 $h_s = M_r h$ 和 $t_s = M_r t$, 然后在语义空间中建模使得 $h_s + r \approx t_s$ 。TransR模型对三元组 (h, r, t) 定义了评分函数, 如公式(1)所示, 其中 $h, t \in \mathbb{R}^n, r \in \mathbb{R}^m$, h, r, t, h_s, t_s 的 L_2 范数均小于等于1。

$$f_r(h, t) = \|h_s + r - t_s\|_{L_1/L_2} \quad (1)$$

4.2 实体描述信息表示

实体文本的描述中包含着十分丰富的信息, 可以作为知识图谱的补充, 将实体的文本描述信息融入到藏语知识图谱表示学习中, 在一定程度上可以缓解藏语知识图谱数据稀疏的问题。我们利用收集的12,449条文本描述信息, 采用了融合多头注意力和关系注意力的Transformer模型, 将文本描述信息与实体关系相关联, 捕捉到更加多元的特征。

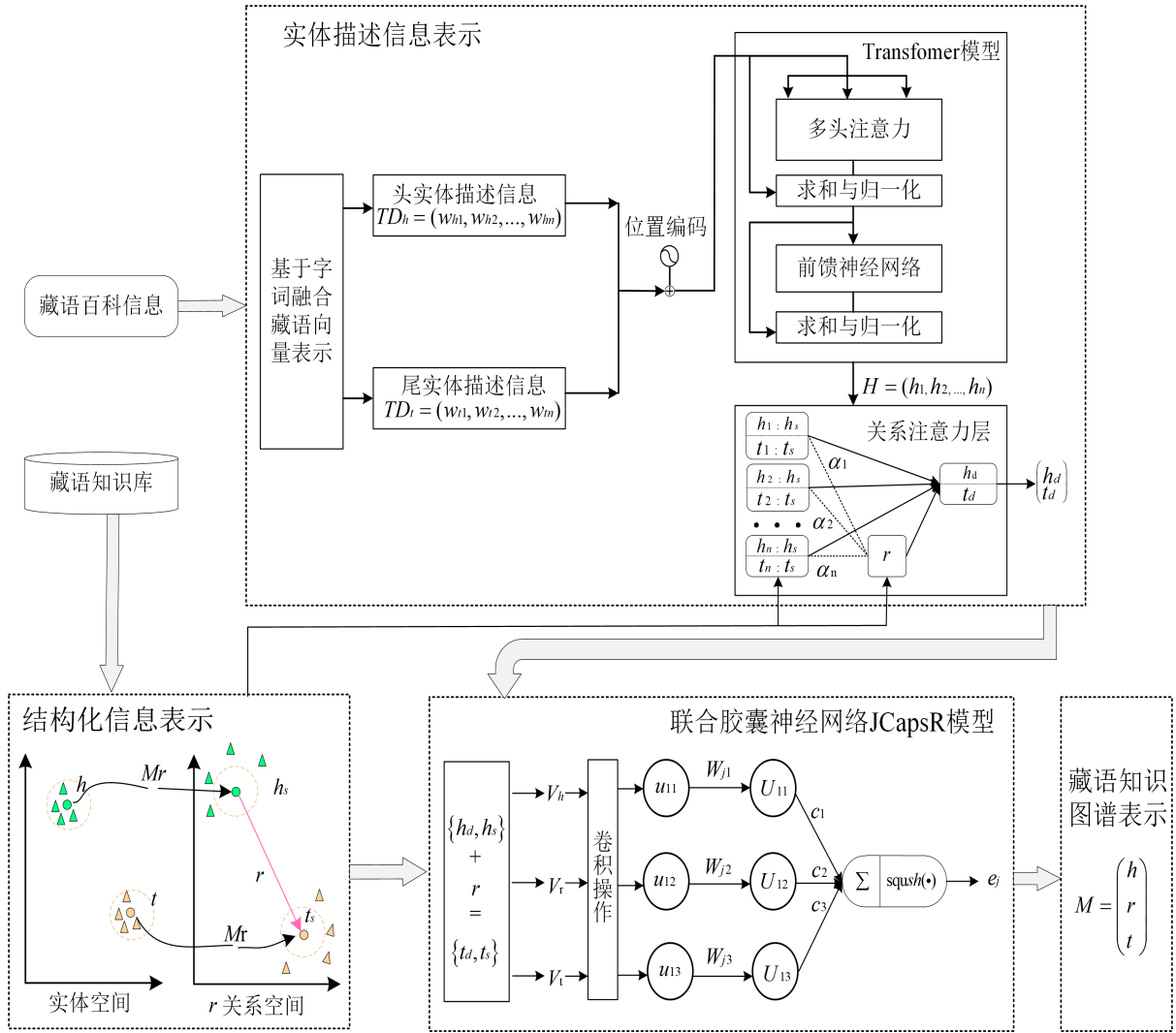


图 1. 藏语知识图谱表示学习框架图

4.2.1 基于字词融合的藏语向量表示

为了获得更丰富的语义信息，本文采用音节和词两种切分方式，分别获得音节和词级别的向量，融合获得藏语的词向量表示，如图2所示。（1）音节级预处理，藏语音节之间有分隔符“.”，本文采用双向长短期记忆网络（BiLSTM）进行编码，得到基于音节切分的实体描述 $(s_1^d, s_2^d, s_3^d, \dots, s_n^d)$ 。（2）词级预处理，我们使用了藏语分词工具进行分词（龙从军et al., 2015），采用预训练模型Glove进行编码，生成维度 $d = 100$ 的向量，得到的每一个实体描述为 $(w_1^d, w_2^d, w_3^d, \dots, w_n^d)$ 。最后将音节向量和词向量通过两层的Highway Network进行融合，得到实体描述 $Desc = \{w_1, w_2, \dots, w_n\}$ ，其中， w_i 表示融合后的词向量。在此基础上我们可以得到头实体描述信息的表示 $TD_h = (w_{h1}, w_{h2}, w_{h3}, \dots, w_{hn})$ ，尾实体描述信息的表示 $TD_t = (w_{t1}, w_{t2}, w_{t3}, \dots, w_{tn})$ 。

例句	ཆ་རིང་ཅི་རིང་དཔེ་མཛད་ཁང་དུ་སློབ་བྱུང་བྱེད་དུ་སྤང་།
	才让今天去图书馆学习了
音节级分词	ཆ་/རིང་/ཅི་/རིང་/དཔེ་/མཛད་/ཁང་/དུ་/སློབ་/བྱུང་/བྱེད་/དུ་/སྤང་།
词级分词	ཆ་རིང་/ཅི་རིང་/དཔེ་མཛད་ཁང་/དུ་/སློབ་བྱུང་/བྱེད་དུ་/སྤང་།

图 2. 藏语音节、词的切分示例

接下来，我们需要对头、尾实体描述信息序列中词汇位置进行编码。Transformer模型采用三角函数计算的方式来获得位置向量，如公式（2）-（3）所示。

$$PE(pos, 2i) = \sin\left(\frac{pos}{1000^{\frac{2i}{d_{model}}}}\right) \quad (2)$$

$$PE(pos, 2i + 1) = \cos\left(\frac{pos}{1000^{\frac{2i}{d_{model}}}}\right) \quad (3)$$

其中， pos 表示单词在句子中的位置， d_{model} 表示位置向量的维度（与词向量 $d = 100$ 相同），公式（2）表示偶数位置的向量表示，公式（3）表示奇数位置的向量表示。使用这种位置编码函数不仅能够很好的表示输入语句的绝对位置和相对位置的关系，还可以处理更长的句子。最后，得到一个包含位置信息的向量 $P = (h_{p1}, h_{p2}, \dots, h_{pn})$ ，其中 h_{pi} 表示的是头（尾）实体描述信息中第 i 个词语的位置嵌入。将词嵌入 TD_h 和位置嵌入 P 连接起来获得输出向量 $S = (h_{s1}, h_{s2}, \dots, h_{sn})$ ，作为Transformer编码层的输入向量。

4.2.2 融合多头注意力和关系注意力的Transformer模型

将输入向量 $S = (h_{s1}, h_{s2}, \dots, h_{sn})$ 输入Transformer模型的编码层，采用多头注意力机制和两层前馈网络，经求和和归一化，得到实体描述信息表示。其中， h_{si} 通过三个给定的参数矩阵 W^Q ， W^K ， W^V 进行线性变换，生成 $Query(Q)$ ， $Key(K)$ ， $Value(V)$ 三个向量。计算如公式（4）所示。

$$Q = SW^Q, K = SW^K, V = SW^V \quad (4)$$

注意力机制聚焦的过程通过函数的权重系数进行体现，权重越大，说明将更多的注意力投射到了对应的值上，Transformer模型中多头注意力的基本操作流程包括并行线性变换，并行计算缩放点积注意力，多头结果拼接以及再次线性变换四个步骤，此过程如公式（5）-（8）所示。

$$Q_i = QW_i^Q, K_i = KW_i^K, V_i = VW_i^V, i = 1, \dots, 8 \quad (5)$$

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (6)$$

$$head_i = Attention(Q_i, K_i, V_i), i = 1, \dots, 8 \quad (7)$$

$$Y = MultiHead(Q, K, V) = Concat(head_1, head_2, \dots, head_h)W^O, h = 8 \quad (8)$$

在本文中，我们设定 $W_i^Q, W_i^K, W_i^V \in \mathbb{R}^{512 \times 64}$ ， $W^O \in \mathbb{R}^{512 \times 512}$ ， $head_i \in \mathbb{R}^{64}$ 。在计算缩放点积注意力时， $d_k = 64$ 是 Q, K 向量的维度，为了防止内积过大，需要除以 $\sqrt{d_k}$ 将方差稳定控制为1，同时还能有效的解决梯度消失的问题。

在完成多头注意力操作之后，为了保持数据特征分布的稳定性，需要对输入数据进行求和与归一化操作。然后，输入到两层前馈网络中，在每一个全连接层中，我们引入Dropout以10%的概率随机删除网络中的一些隐藏神经元，保持输入输出神经元不变，这样提高了网络的稳定性。最后，输出每个藏语头（尾）实体文本描述信息的向量表示 $H = (h_1, h_2, \dots, h_n)$ 。

通过Transformer模型的编码层获得了 $H = (h_1, h_2, \dots, h_n)$ ，该向量只包含文本描述的信息，没有考虑到结构化的三元组信息，尤其是没有考虑到结构化信息的关系信息。因此，本文在Transformer模型之后引入关系注意力机制，获取描述信息中与结构化相关的信息。

我们使用一个全连接神经网络来计算每个文本描述信息单词的权重，输入是结构化三元组的头(尾)实体和关系，以及文本描述信息的单词。输出得到与结构化三元组关系更为接近的文本描述信息的单词表示。其中，注意力机制的计算如公式（9）-（11）所示。

$$f_i = W_a[h_s; h_i] + Er \quad (9)$$

$$\alpha_i = \text{softmax}(f_i) = \frac{\exp(f_i)}{\sum_{j=1}^N \exp(f_j)} \quad (10)$$

$$h_d = \sum_{i=1}^N \alpha_i h_i \quad (11)$$

W_a 是权重矩阵, E 是关系 r 的参数, 将公式 (9) 的结果 f_i 做归一化操作, 得到的数值与文本描述信息的单词加权求和, 最终获得文本描述信息的向量表示 h_d 。

4.3 联合胶囊神经网络JCapsR模型

本文对藏语文本描述信息(h_d, t_d)和三元组结构信息(h_s, t_s)进行融合并联合训练, 将实体的两种不同类型的向量映射到同一语义空间中, 共享相同的关系向量。能量函数如公式 (12) 所示。

$$E(h, r, t) = \alpha_1 \|h_s + r - t_s\| + \alpha_2 \|h_s + r - t_d\| + \alpha_3 \|h_d + r - t_s\| + \alpha_4 \|h_d + r - t_d\| \quad (12)$$

其中, $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ 是控制各项函数的超参数。融合后得到三元组向量 V_h, V_r, V_t , 如公式 (13) - (15) 所示, 其中 W_s 和 W_d 是共享参数, \odot 为哈达玛积是同阶的矩阵乘法。

$$V_h = W_s \odot h_s + W_d \odot h_d \quad (13)$$

$$V_r = r \quad (14)$$

$$V_t = W_s \odot t_s + W_d \odot t_d \quad (15)$$

我们将融合后的三元组向量作为矩阵 $A = (V_h, V_r, V_t) \in \mathbb{R}^{k \times 3}$ 输入到胶囊神经网络模型中, 其中 A_i 表示矩阵 A 的第 i 行, k 是词向量的维度为100。此胶囊神经网络模型主要分为两部分: 卷积层和胶囊层。

(1) 卷积层

将矩阵 A 作为卷积层的输入, 采用三个不同的窗口, 其大小为 $j \in (1, 2, 3)$, 每个窗口使用2个卷积核 ω_j , 对矩阵 A 进行卷积操作, 对于每一个卷积核, 通过公式 (16) - (18) 进行运算。

$$\mu_{j,i} = \xi(\omega_j \cdot A_i, : + b) = \text{ReLU}(\omega_j \cdot A_i, : + b) \quad (16)$$

$$q_j = [\mu_{j1}, \mu_{j2}, \dots, \mu_{jk}] \in \mathbb{R}^k \quad (17)$$

$$q = [q_1, q_2, q_3] \quad (18)$$

其中, “ \cdot ”是点积运算, $b \in \mathbb{R}$ 是偏置项, ξ 是激活函数ReLU, $\mu_{j,i}$ 表示卷积核 ω_j 在矩阵 A 的第 i 行上步长为1的每个位置的特征元素, q_j 是不同窗口卷积操作下获得的特征图, q 是卷积操作获得的特征图集合。三元组向量通过卷积层会产生6个特征图作为第一个胶囊层的输入。

(2) 胶囊层

胶囊模型中使用了两个胶囊层, 第一层为每个特征图列表构造了 k 个胶囊。我们将特征图集合中相同维度的特征封装到一起, 用来捕获三元组向量中不同位置的特征。对于每一个胶囊, 都会产生相应的向量 u_{ji} , 然后将向量 u_{ji} 乘以权重矩阵 W_{ji} 获得向量 U_{ji} , 乘以耦合系数 c_i , 对向量 U_{ji} 加权求和, 得到第二层胶囊的输入向量 s_j 。最后, 对向量 s_j 执行非线性压缩函数 $\text{squash}()$, 产生一个输出向量 e_j , 向量 e_1, e_2, e_3 经过加权求和后获得向量 e , 其长度表示三元组的分数。上述过程如公式 (19) - (21) 所示。

$$e_j = \text{squash}(s_j), s_j = \sum_i c_i U_{ji} = \sum_i c_i W_{ji} u_{ji} \quad (19)$$

$$squash(s_i) = \frac{\|s_i\|^2}{1 + \|s_i\|^2} \frac{s_i}{\|s_i\|} \quad (20)$$

$$e = \sum_{j \in \{1,2,3\}} e_j \quad (21)$$

4.4 模型训练

最终，获得的三元组表示 $M = (h, r, t)$ ，如公式 (22) 所示。其中， $f(h, r, t)$ 是得分函数， Ω 表示卷积层滤波器共享参数， $*$ 表示卷积操作， $JCapsR$ 表示联合胶囊神经网络操作， $ReLU$ 表示激活函数。

$$f(h, r, t) = \| JCapsR(ReLU([v_s, v_r, v_o] * \Omega)) \| \quad (22)$$

在训练的过程中，将三元组的正例随机替换任意一元素后得到三元组的负例，随机生成负例并使用最大间隔方法来提高训练的速度并改善词向量的质量。损失函数如公式 (23) 所示。

$$L = \sum_{(h,r,t) \in T} \sum_{(h',r',t') \in T'} \alpha \max(\gamma + E(h, r, t) - E(h', r', t'), 0) + \beta \log[(1 + \exp(-f(h, r, t)))(1 + \exp(f(h, r, t)))] \quad (23)$$

其中， T 、 T' 分别表示三元组正例集合和负例集合， α 、 β 是超参， $E(h, r, t)$ 是公式 (12) 中的能量函数， γ 表示正负样本之间的间距。在随机生成负例时，为了避免随机替换生成真实的三元组，我们可以对其添加条件： $\forall (h', r', t' \in T')$ ， $h', r', t' \notin T$ ，可以使得训练模型具有更强的泛化能力。

5 实验结果与分析

5.1 数据集

本文采用藏语数据集 TD12K-5 (Sun et al., 2021) 进行了实验。在 TD12K-5 中，包括了 50,000 个藏语三元组，40,750 个实体，2,581 个关系，12,449 个文本描述信息。平均每个实体包含 5 个三元组，一个实体对应的三元组的最大数值为 191，最小数量为 2。

5.2 评估指标

本文采用 Mean Rank 和 Hit@10 作为测评指标 (Bordes et al., 2013)。

(1) Mean Rank: 对正确实体的平均排名。对于每个测试集，我们会将头实体用其他的实体进行替换，将被替换的实体称之为被破坏的实体，将这些被破坏的实体放到模型中进行预测，按照升序对正确预测的实体进行排序。此测评指标表示所有测试实例中正确答案的平均排序，Mean Rank 值越小，证明模型的实测效果越好。

模型	数据集	Mean Rank		Hits@10	
		Raw	Filter	Raw	Filter
CNN+TransE	TD12K-5	408	406	70%	73%
CNN+TransR	TD12K-5	407	404	71%	72%
CNN+ATT+TransE	TD12K-5	382	380	74%	78%
CNN+ATT+TransR	TD12K-5	380	378	75%	78%
Transformer+TransR	TD12K-5	405	402	72%	75%
Transformer+ATT+TransR	TD12K-5	375	374	79%	81%
Transformer+ATT+TransR+JCapsR	TD12K-5	370	367	82%	83%
Joint Model (Sun et al., 2021)	TD12K-5	376	371	79%	80%

表 1. 不同模型的实验结果

(2) Hits@N: N命中率。一般选择 $N = 10$, 表示排名前10的答案中有正确实例的比例。比例越大时, 模型能够正确表示三元组关系的能力越强。在某些情况下, 替换的实体可能会成为训练集和验证集中的正确三元组, 这样会干扰实验的结果, 因此, 我们需要处理完测试集后, 及时的删掉已经在训练集和验证集中出现的三元组, 处理后的测试集称为“Filter”, 不做替换处理的测试集称为“Raw”。

5.3 实验结果

采用TD12K-5数据集, 我们进行了如下实验, 实验结果如表1和图3、4所示。

(1) **CNN+TransE**: 采用CNN处理实体描述信息, TransE处理三元组结构化信息, 将结构化信息和实体文本描述信息进行融合, 此模型作为本文的基线系统。

(2) **CNN+TransR**: 采用CNN处理实体描述信息, TransR处理三元组结构化信息。

(3) **CNN+Attention+TransE**: 采用CNN处理实体描述信息, TransE处理三元组结构化信息, 引入Attention来聚焦关系语义特征。

(4) **CNN+Attention+TransR**: 采用CNN处理实体描述信息, TransR处理三元组结构化信息, 引入Attention来聚焦关系语义特征。

(5) **Transformer+TransR**: 在TransR模型的基础上, 用Transformer模型表示实体描述信息。

(6) **Transformer+Attention+TransR**: 引入了多头注意力机制和关系注意力机制。

(7) **Transformer+Attention+TransR+JCapsR**: 引入联合胶囊神经网络JCapsR模型对三元组特征提取进行优化。

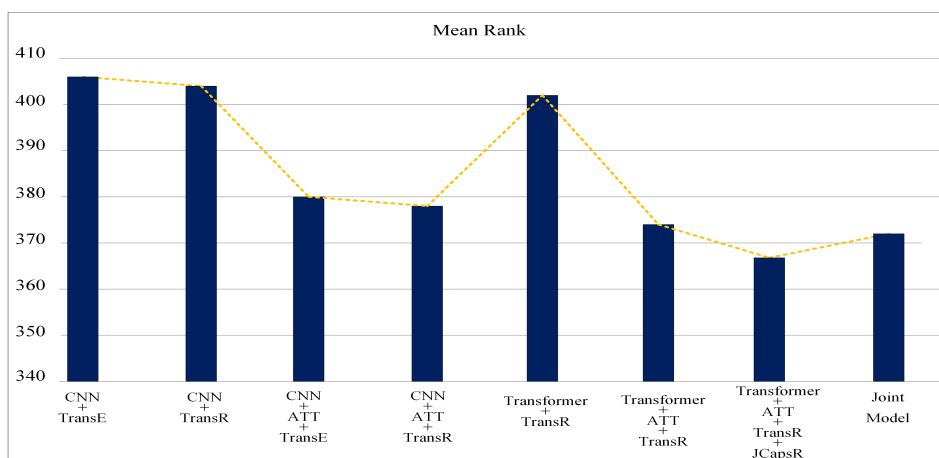


图 3. 不同模型在Mean Rank上的实验结果

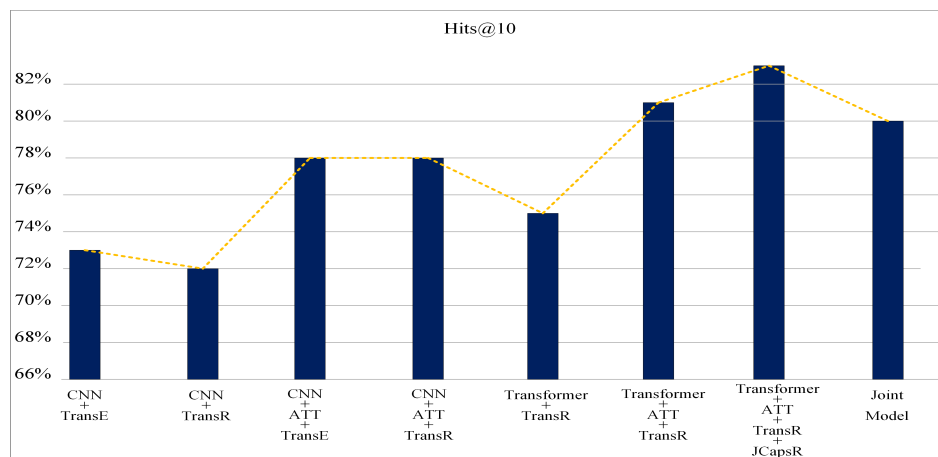


图 4. 不同模型在Hits@10上的实验结果

对上述实验结果进行分析, 可以得出如下结论:

(1) 与基线系统CNN+TransE比较, CNN+TransR模型的效果有所提升, 而加入注意力机制, CNN+ATT+TransE模型的Hits@10值提高了4%, CNN+ATT+TransR模型的Hits@10值提高了5%。这表明, TransR模型在对藏语知识图谱三元组结构化信息表示上的表现要比TransE模型要好, TransR模型将实体空间和关系空间进行分离, 每个实体和关系映射到不同的语义空间上训练表征向量, 在一定程度上解决了藏语语义歧义的问题。此外, 注意力机制的加入对模型的效果提升明显, 对于数据相对稀疏的藏语知识库, 注意力机制可以从实体文本描述信息中获得更有效的信息, 提高了知识图谱的表示能力。

(2) 采用Transformer模型和Transformer+Attention模型学习藏语实体描述信息表示, 相比与基线系统, Hits@10值分别提高了2%和9%, 说明Transformer的效果优于CNN, 多头注意力机制和位置编码以及关系注意力机制的引入, 在藏语实体描述信息表示中, 提升了注意力层的性能, 扩展了关注不同位置的能力, 增强了语义特征提取能力和长距离特征捕获能力, 从而可以关注到更多元的信息。

(3) 本文提出的联合胶囊神经网络JCapsR模型, 相比基线系统, Hit@10值提高了12%, Mean Rank值下降了39, 在所有模型中取得了最好的结果, 这说明胶囊神经网络在特征提取方面比CNN的效果要好, 胶囊神经网络使用向量作为模型的输入、输出, 更加准确的表征藏语三元组中实体和关系, 从而能够更好的建模神经网络中内部知识表示的分层关系。

(4) 将本文提出的模型与Joint Model (Sun et al., 2021)的实验结果对比, 发现JCapsR模型的Mean Rank值均比Joint Model要低, 并且Hit@10值提升了3%, 证明在藏语知识图谱表示学习方面, JCapsR模型的性能要优于Joint Model, 可以说明Transformer模型在藏语实体描述信息表示方面的效果明显优于CNN, 另外胶囊神经网络在文本处理任务中表现出了更强的特征提取能力。

6 总结与展望

本文提出了一种用于藏语知识图谱表示的联合胶囊神经网络JCapsR模型。采用TransR模型生成藏语知识图谱的结构化信息表示, 使用融合多头注意力和关系注意力的Transformer模型表示藏语实体的文本描述信息。最后, 采用JCapsR进一步提取三元组在知识图谱语义空间中的关系。实验结果表明, 相比基线系统, 联合胶囊神经网络JCapsR模型提高了藏语知识图谱表示学习的效果。未来我们将尝试对低资源语言知识图谱表示进行优化, 并进一步扩充知识库的规模, 尝试引入更多的外部辅助信息来实现多模态嵌入, 达到增强知识表示学习的效果。

致谢

本论文得到了国家自然科学基金项目(61972436)的资助。

参考文献

- Bordes, A, Usunier, M, and Garcia Duran, A. 2013. Translating embeddings for modeling multi-relational data. *The Neural Information Processing Systems Foundation*, 2:2787–2795.
- 刘知远, 孙茂松, 林衍凯, 谢若冰. 2016. 知识表示学习研究进展. *计算机研究与发展*, 53(2):247–261.
- Wang, Z, Zhang, J and Feng, J. 2014. Knowledge graph embedding by translating on hyperplanes. *Proceedings of the AAAI Conference on Artificial Intelligence*, 28(1).
- Ji, G, He, S and Xu, L. 2015. Knowledge graph embedding via dynamic mapping matrix. *Proceedings of the 53rd annual meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, 1:687–696.
- Lin, Y, Liu, Z, Sun, M and Liu, Y, Zhu, X. 2015. Learning entity and relation embeddings for knowledge graph completion. *Proceedings of the 29th AAAI Conference on Artificial Intelligence*, 29(1):2181–2187.
- Xiao, H, Huang, M, and Hao, Y. 2015. TransA: an adaptive approach for knowledge graph embedding. *arXiv preprint*, arXiv:1509.05490.
- 朱臻, 孙媛. 2015. 基于SVM和泛化模板协作的藏语人物属性抽取. *中文信息学报*, 29(6):220–227.

- Vaswani. A, Shazeer. N, and Parmar. N. 2017. Attention is all you need. *Advances in Neural Information Processing Systems*, 6000–6010.
- 冯小兰, 赵小兵. 2019. 汉藏双语旅游领域知识图谱系统构建. *中文信息学报*, 33(11):66–72.
- Wang. Z, Zhang. J, Feng. J and Chen. Z. 2014. Knowledge graph embedding by translating on hyperplanes. *Proceedings of the 28th AAAI Conference on Artificial Intelligence*, 28(1):1112–1119.
- 夏天赐, 孙媛. 2018. 基于联合模型的藏文实体关系抽取方法研究. *中文信息学报*, 32(12):76–83.
- Vrandečić. D and Krötzsch. M. 2014. Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 57(10):78–85
- Auer. S, Bizer. C, and Kobilarov. G. 2007. Dbpedia: a nucleus for a web of open data. *Proceedings of the 6th International Semantic Web Conference*, 722–735.
- 王丽客, 孙媛, 夏天赐. 2020. 基于远程监督的藏文实体关系抽取. *中文信息学报*, 34(03):72–79.
- Bollacker. C, Evans. C, and Paritosh. P. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. *Eurographics Association*, 57(10):78–85.
- Bordes. A, Glorot. X, and Weston. P. 2012. Joint learning of words and meaning representations for open-Text semantic parsing. *Mechine Learning*, 94(2):233–259.
- Lin. Y, Liu. Z, and Sun. M. 2015. Learning entity and relation embeddings for knowledge graph completion. *Proceedings of the AAAI Conference on Artificial Intelligence*, 29(1)
- Sabour. S, Frosst. N, and Hinton. G. E. 2017. Dynamic routing between capsules. *arXiv preprint*, arXiv:1710.09829.
- Hinton, Geoffrey. E, and Nicholas. Frosst. 2018. Matrix capsules with EM routing. *ICLR'2018 6th International Conference on Learning Representations*, 1–15.
- Sun. Y, Chen A, Chen. C, Z, Xia. T and Zhao. X. 2021. A joint model for representation learning of Tibetan knowledge graph based on Encyclopedia. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 20(2):1–17.
- Wang. C, Ju. S, and Sun. J. 2020. Incorporating Multi-scale Feature Attention into Capsule Network and its Application in Text Classification. *Proceedings of the 19th Chinese National Conference on Computational Linguistics*, 719–730.
- Cheng. J, Zhang. F, and Yang. Z. 2020. Knowledge Graph Representation Learning With Multi-Scale Capsule-Based Embedding Model Incorporating Entity Descriptions. *IEEE Access*, 1–15.
- 龙从军, 刘汇丹, 诺明花, 吴健. 2015. 基于藏语字性标注的词性预测研究. *中文信息学报*, 29(5):211–216.