DOI:10.13718/j.cnki.xsxb.2021.01.004

分布式语义框架在自然语言理解中的应用①

李潇雯[1], 朱齐亮[2]

1. 山西工商学院 计算机信息工程学院, 太原 030006; 2. 华北水利水电大学 信息工程学院, 郑州 450045

摘要:为了学习非结构化文本与对应的结构化语义知识之间的嵌入语义对应关系,本文提出了一种用于自然语言理解(Natural Language Understanding, NLU)的分布式语义向量学习框架. 该语义框架使用长短期记忆对输入序列进行编码以生成文本向量,然后将意图标签、时隙标记和时隙值向量合并生成分布式语义向量,通过最小化文本输出向量与语义框架向量的距离,将语义等价向量放置在向量空间中,最后采用意图重构和时隙标签生成损失作为目标得分以学习鲁棒的语义向量. 实验结果表明,所学习的语义向量包含语义信息,该语义框架在 NLU 结果重新排列方面均优于测试的 NLU 系统.

关 键 词:自然语言理解;分布式表示;语义向量学习;语义框架重构

中图分类号:TP391 **文献标志码**:A **文章编号**:1000-5471(2021)01-0019-06

计算机科学的不断发展和成熟对人机界面的智能化提出了新的要求[1-2]. 机器的智能化导致对语言文字的处理深度和广度越来越高,在界面层,识别、理解和翻译自然语言是最重要的要求之一[3-4]. 自然语言理解(Natural Language Understanding, NLU)是实现聊天机器人、移动秘书和智能扬声器等自然用户界面的核心技术,自然语言理解的目标是从自然语言中提取意义并推断用户意图. NLU 嵌入模型有助于分析理解非结构化文本和与其对应的结构化语义知识之间的关系,对 NLU 的研究者和实践者都是必不可少的[5-6].

从人工智能的角度看,NLU 的任务是建立一种能够给出像人那样理解、分析并回答自然语言结果的计算机模型[7-8]. NLU 通常涉及 2 个任务:识别用户意图和提取特定领域的实体,其中,识别用户意图一般表述为句子分类,需对每个句子完成单个或多个意图标签的预测[9-10]. 提取特定领域实体通常被称为时隙填充,其中仅有部分句子被提取并用领域实体进行标记[11],NLU 使用统计建模来完成意图识别和时隙填充任务.

为了使 NLU 技术在实践和科学研究中发挥最大的作用,文献[12]提出了一种用于自然语言理解的多任务基深层神经网络,该方法将多任务学习与预训练语言模型相结合进行语言表示学习,利用大量的跨任务数据和正则化效应来适应新的任务和领域,在领域适应实验中具有优异的泛化能力. 文献[13]采用 Transformers 的双向编码器表示(Bidirectional Encoder Representations from Transformers, BERT)顺序推荐模型,用深层次的双向自编码建模用户行为序列. 文献[14]提出了一种基于序列到序列模型和指针网络的生成性时隙填充神经网络模型,用来预测只有句子级语义标注时对话数据的时隙值. 该模型通过在复制和生成未登录词(out-of-vocabulary, OOV)之间切换,可以绕过单词级标注的需要,并克服在实际自然语言处理中常见的 OOV 问题.

① 收稿日期:2020-04-30

基金项目:教育部产学合作协同育人项目(201901195002);山西省教育科学"十三五"规划 2020 年度"互联网+教育"专项课题(HLW-20141);山西省 1331 工程立德树人好老师课程建设计划支持人选项目(201832);山西工商学院 2019 年"助推计划"专项课题(201953).

作者简介:李潇雯,硕士,讲师,主要从事数据挖掘、大数据等方面的研究.

上述NLU文献使用统计建模来完成意图识别和时隙填充任务以及输入表示，但均未同时将文本和语义框架表示为矢量形式，因此，本文提出了一种同时学习文本和语义框架的分布式语义向量学习框架，在单个框架中同时执行原始文本到向量和结构化文本到向量的方法，以便更直接地学习语义表示. 在该框架中，通过最小化文本输出的语义向量与语义框架阅读器之间的距离，将文本和语义框架分别投影到向量空间. 语义框架重构技术用于在嵌入向量与其对应的语义框架之间导出一对一的映射. 为了学习语义框架表示的鲁棒向量形式，采用文本投影向量与语义框架间的对应关系作为目标得分，并采用重构意图和标签生成损失作为目标得分. 实验结果表明，本文所提出的语义框架实现了自然语言库的可视化，可以对多个系统的自然语言库结果进行重新排序.

1 分布式语义向量学习框架

本文分布式语义向量学习框架由文本阅读器、语义框架阅读器和语义框架编写器组成. 文本阅读器将标记序列嵌入到分布式向量表示中，语义框架阅读器读取结构化文本并将每个文本编码为一个向量，语义框架编写器根据向量表示生成一个符号语义框架. $\boldsymbol{v}_t$ 表示从文本阅读器派生的向量语义框架，$\boldsymbol{v}_s$ 表示从语义框架阅读器派生的向量语义框架.

设一对对应的文本和语义框架(t, s)在原始文本表示形式(x_T)中具有相同的语义，并且语义框架表示形式(x_S)可以编码为共享嵌入向量空间 Z 中的向量 $\boldsymbol{v}$. 文本阅读器函数 R_T 读取原始文本并将每个原始文本编码为向量，语义阅读器函数 R_S 将结构化文本编码为语义向量，W 是将语义向量解码为符号语义框架的写函数. 语义框架和语义框架的向量形式之间的关系如图1所示.

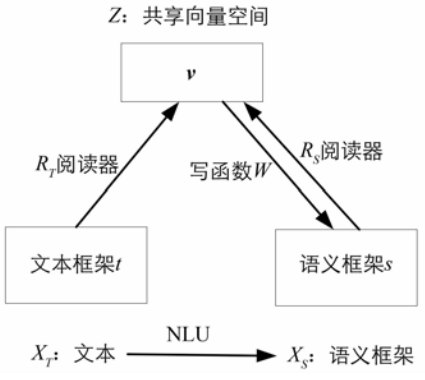

图1 语义向量学习框架

1.1 输入序列编码

文本阅读器 R_T 读取一系列输入标记并将每个标签编码为一个向量 $\boldsymbol{v}_t$，实现了神经语句编码器. 本文使用长短期记忆(Long Short-Term Memory，LSTM)对输入序列进行编码. 编码过程可以定义为

$$\vec{h}_s = R_{\text{text}}(E_X(x_S), \vec{h}_{s-1}) \tag{1}$$

$$\boldsymbol{v}_t = \sigma(\vec{h}_s) \tag{2}$$

其中，$s = \{1, 2, \cdots, S\}$，$\vec{h}_s$ 为时间 s 上输入序列上的前向隐藏状态，R_{text} 是循环神经网络(Recurrent Neural Network，RNN)单元，E_X 是输入文本标记 x 的嵌入函数，$E_X(x_S)$ 是标记嵌入函数，它表示时间 s 返回标签 x 的分布式矢量，$\sigma(z)$ 是sigmoid函数. 最后RNN的输出 $\vec{h}_s$ 表示为 $\boldsymbol{v}_t$，它是从文本中派生出来的语义向量.

1.2 构建分布式语义框架

本文使用语义框架阅读器来构建分布式语义框架，它由意图标签、时隙标记和时隙值等结构化标签组成. 本文将意图标签作为一个符号来处理，时隙标记和时隙值被当作一系列符号来处理. 例如，“请列出星期一从北京到上海的所有航班”这句话的处理方式如下：

意图标签：航班

时隙标记序列：[出发城市，终点城市，出发日期]

时隙值序列：[北京，上海，星期一]

意图阅读器是一个简单的嵌入函数 $v_{\text{intent}} = E_I(i)$，其中 E_I 为意图标签的嵌入函数，它返回一个句子的意图标签 i 的分布式矢量表示.

堆栈LSTM层用于读取时隙标记和时隙值的序列. 设 $E_S(o)$ 是一个以 o 为标志的时隙标签的嵌入函数，

$E_V(a)$ 是一个以 a 为标志的时隙值的嵌入函数. 嵌入结果 $E_S(o_m)$ 和 $E_V(a_m)$ 在时间步 m 处串联，合并后的向量分别在每个时间步长处都反馈到叠加层. 时隙标记和时隙值序列的读取结果从 M 的循环神经网络单元 RNN 的最终输出结果中提取，其中，M 为句子中的时隙数. 最后，构造出如下的分布式语义框架：

$$\boldsymbol{v}_s = \sigma(W_{sf}([\boldsymbol{v}_{\text{intent}}; \boldsymbol{v}_{\text{tag, value}}]) + b_{sf}) \tag{3}$$

其中[；]表示向量连接运算符，$\boldsymbol{v}_{\text{intent}}$ 表示意图识别输出，$\boldsymbol{v}_{\text{tag, value}}$ 表示时隙标签和时隙值的阅读器输出，$\boldsymbol{v}_s$ 的维度和 $\boldsymbol{v}_t$ 相同. 所有的嵌入权重均随机初始化并通过训练过程学习.

1.3　语义框架编写器和损失函数

本文语义框架编写器从语义上学习文本的合理矢量表示形式和相关的语义框架. 首先通过最小化文本输出的语义向量 $\boldsymbol{v}_s$ 和语义框架阅读器 $\boldsymbol{v}_t$ 之间的距离，将语义等价向量放置在向量空间中. 然后根据期望语义向量的属性计算相应的损失函数.

本文期望语义向量的 2 个属性：嵌入对应和语义框架重构.

嵌入对应的损失：嵌入语义上有意义的量，并在向量空间中计算它们的距离，距离损失度量了向量空间中来自文本阅读器与来自语义框架阅读器的语义向量之间的相似性. 损失函数定义为：

$$L_{\text{dist}} = D_{\text{dist}}(\boldsymbol{v}_t, \boldsymbol{v}_s) \tag{4}$$

其中，D_{dist} 函数可以是任何矢量距离度量，本文采用欧几里得距离和余弦距离.

语义框架重构的损失：语义框架重构用于得出嵌入向量及其对应的语义框架之间的一对一映射. 内容损失提供了一种语义框架向量包含多少语义信息的方法. 在没有内容损失的情况下，$\boldsymbol{v}_t$ 和 $\boldsymbol{v}_s$ 趋向于迅速收敛到零向量，这意味着无法学习语义表示. 为了度量内容的保持性，从语义向量中生成符号语义框架，并计算原始语义框架与生成语义框架之间的差异.

由于语义框架的时隙值需要较大的词汇量来生成时隙值，因此设计了一个简化的语义框架来解决时隙值的生成问题. 本文通过简单地从相应的语义框架中删除时隙值，创建一个简化的语义框架，在这个简化的语义框架上执行内容丢失计算. 使用简化语义框架的另一个优点是，所学习的分布式语义向量对词汇的敏感度较低，因此具有更高的抽象能力.

对于内容丢失，将测量意图标签和时隙标签的生成质量. 意图生成网络可以简单地用线性投影定义为如下所示：

$$\boldsymbol{y}_{\text{intent}} = W'_I \boldsymbol{v} + b_1 \tag{5}$$

其中，$\boldsymbol{v}$ 是语义向量，W_I 为意图写入函数，$\boldsymbol{y}_{\text{intent}}$ 是输出向量，b_1 为偏置量.

时隙标记生成网络定义为

$$\vec{q}_m = R_G(\boldsymbol{v}, \vec{q}_{m-1}) \tag{6}$$

$$y^m_{\text{slot}} = W'_S \vec{q}_m + b_S \tag{7}$$

其中，R_G 是 RNN 单元，b_S 为偏置量. 语义向量 $\boldsymbol{v}$ 被复制并重复地输入到每个 RNN 输入中. RNN 的输出用于 W'_S 投影到时隙标签空间.

生成的标签向量和参考标签向量之间的交叉熵可以定义为

$$L_{\text{intent}} = H(\hat{\boldsymbol{y}}_{\text{intent}}, \boldsymbol{y}_{\text{intent}}) \tag{8}$$

$$L_{\text{slot}} = \frac{1}{M}\sum_{m=1}^{M} H(\hat{\boldsymbol{y}}^m_{\text{slot}}, \boldsymbol{y}^m_{\text{slot}}) \tag{9}$$

其中，M 为句子中的时隙数(包括时隙结束序列符号)，$H(p, q)$ 为 CrossEntropy 交叉熵函数.

结合意图和时隙损失，从语义向量 $\boldsymbol{v}$ 重构语义框架的内容损失(L_{content})，可以定义为

$$L_{\text{content}} = \begin{cases} 0, & \text{无内容丢失} \\ L_{\text{intent}}, & \text{仅考虑意图} \\ L_{\text{slot}}, & \text{仅考虑时隙} \\ L_{\text{intent}} + L_{\text{slot}}, & \text{考虑意图和时隙} \end{cases} \tag{10}$$

在上述描述中，可以通过进一步的处理从 $\boldsymbol{v}_t$ 和 $\boldsymbol{v}_s$ 中选择分布式语义向量 $\boldsymbol{v}$，则

$$\boldsymbol{v}=\begin{cases}\boldsymbol{v}_t & T:\text{text}\\ \boldsymbol{v}_s & S:\text{semantic frame}\\ 0.5*(\boldsymbol{v}_t+\boldsymbol{v}_s) & A:\text{average}\\ [\boldsymbol{v}_t;\ \boldsymbol{v}_s] & C:\text{concat}\end{cases} \tag{11}$$

最后，用于学习语义框架表示的总损失值(L)定义为

$$L=L_{\text{dist}}+L_{\text{content}}+L_{\text{slot}} \tag{12}$$

1.4 多种形式的距离测量

使用所学的文本和语义框架阅读器，不仅可以测量来自同一表示(文本或语义框架)的实例，还可以测量来自不同表示的实例. 将文本表示为 t 和语义框架表示为 s，将文本和语义框架阅读器表示为 R_T 和 R_S. 它们之间的距离测量可按如下方式进行：

$$D_{\text{dist}}(\boldsymbol{v}_t^i,\ \boldsymbol{v}_t^j):\ t_i\rightarrow R_T(t_i)=\boldsymbol{v}_t^i,\ t_j\rightarrow R_T(t_j)=\boldsymbol{v}_t^j \tag{13}$$

$$D_{\text{dist}}(\boldsymbol{v}_t^i,\ \boldsymbol{v}_s^j):\ t_i\rightarrow R_T(t_i)=\boldsymbol{v}_t^i,\ s_j\rightarrow R_S(t_j)=\boldsymbol{v}_s^j \tag{14}$$

$$D_{\text{dist}}(\boldsymbol{v}_s^i,\ \boldsymbol{v}_s^j):\ s_i\rightarrow R_S(s_i)=\boldsymbol{v}_s^i,\ s_j\rightarrow R_S(s_j)=\boldsymbol{v}_s^j \tag{15}$$

在所提出的语义框架中，具有不同形式(文本或语义框架)的实例可以直接在语义向量空间上进行比较.

2 实验结果与分析

为了对所提出的语义框架进行性能评估，本文使用 ATIS2 数据集进行训练和测试(见表 1). ATIS2 数据集由一个带注释的意图和空乘信息搜索任务的时隙语料库组成，ATIS2 数据集带有一个常用的训练和测试拆分. 在本次实验分析中，进一步将训练集进行划分，实验使用的训练集占原训练样本的 90%，开发集占 10%.

表 1 ATIS2 数据集

	训练集	开发集	测试集
句子	4 478	499	882
意向标签	21	15	16
时隙标签	120	96	95
每句话的平均时隙	3.32	3.42	3.17
独特的语言	867	463	440
每句话的平均字数	11.28	11.40	10.23

基于相似度的分类器通过测试样本和标记训练样本之间的相似度，以及训练样本之间的成对相似度，从而得到测试样本的类别标签. 计算每个训练句子的文本语义向量，并用相应的意图标记索引. 当句子被赋予 NLU 系统时，文本阅读器读取句子并生成 $\boldsymbol{v}_t$. 然后，意图标记随距离分数进行升序排列.

表 2 基于相似度的意图分类结果

K	T	S	A	C
1	97.73	77.66	83.45	78.12
3	97.62	86.17	84.24	86.62
5	97.62	86.85	85.60	97.41
10	97.17	87.19	85.71	86.96
40	92.18	83.11	85.60	86.28

表 2 给出了基于相似度的意图分类结果，K 在 1～40 之间变化. 从意图标志的顶部 K 列表中，选择出现最频繁的意图标记作为给定句子的意图标记. 可以看出，所提语义框架具有较好的分类性能，在 $K=1$ 时意图分类性能最优.

对多个 NLU 模块输出结果重新排列是困难的，但这种排列对于构建健壮的 NLU 系统来说是非常重要的. 典型的选择是将结果与每个系统产生的分数进行比较，但是这技术并不总是可行的，语义框架的向量形式为重排序问题提供了一种非常清晰、自然的解决方法. 本文根据对应的 $\boldsymbol{v}_s$ 到 $\boldsymbol{v}_t$ 的距离，重新排序来

自多个 NLU 系统的 NLU 结果(语义框架).

表 3　文本合成法重新排序的性能结果

NLU 系统	意图	时隙		
	Acc	Pre	Rec	F-m
随机	92.86	93.96	92.29	93.12
多数投票重排序	94.10	95.33	93.68	94.64
NLU 得分重排序	95.58	94.81	93.89	94.35
欧几里得重排序	97.05	93.74	91.96	92.84
所提方法重排序	97.05	95.40	94.11	94.75

表 3 给出了文本合成法重新排序算法的性能结果，其中 Acc 表示准确度，Pre 表示精确度，Rec 表示召回率，F-m 表示 F 测量. NLU 结果重新排序的典型选择是多数投票和基于 NLU 分数的排序，多数投票法选择 NLU 系统预测最多的语义框架. 可以看出所提出的基于距离的语义向量重排序方法在意图和时隙嵌入方面都表现出了优越的选择性能，这是因为本文使用学习的语义向量，通过比较文本和语义框架的语义向量值来实现对多个 NLU 系统的重新排序.

表 4　NLU 性能与 v_t 到 v_s 距离的相关性分析

分析方法	意图	时隙			联合
	Acc	Prec	Rec	F-m	Acc
皮尔逊	−0.40	−0.32	−0.33	−0.33	−0.47
斯皮尔曼	−0.47	−0.47	−0.47	−0.47	−0.47

所提出的重排序算法基于以下假设：NLU 系统的质量与 $\boldsymbol{v}_t$ 到 $\boldsymbol{v}_s$ 距离之间存在很强的相关性. 表 4 给出了 NLU 系统的所有测试语句(11 * 882＝9 702 语句)的相关性分析结果. 所有的性能指标(特别是联合指标)均显示 p 值接近零的强相关性(负相关性)，它表明 $\boldsymbol{v}_t$ 到 $\boldsymbol{v}_s$ 的距离越小，NLU 性能越好.

3　结　　论

为了得到有效和有意义的分布式语义表示模式，本文提出了一种用于自然语言理解的分布式语义向量学习框架. 该框架使用深度 LSTM 对输入序列进行编码，之后通过多维度标签组构建分布式语义框架，然后设计语义框架编写器和损失函数以学习鲁棒的语义向量，最后通过多形式距离测量，使具有不同形式(文本或语义框架)的实例可以直接在语义向量空间上进行比较. 实验结果表明，本文提出的分布式语义向量学习框架能够学习文本与提取的语义知识之间的嵌入语义对应关系，同时，该框架在 NLU 输出结果重新排列方面，性能优于测试的 NLU 系统.

参考文献：

[1] BUCHLAK Q D, ESMAILI N, LEVEQUE J C, et al. Machine Learning Applications to Clinical Decision Support in Neurosurgery: an Artificial Intelligence Augmented Systematic Review [J]. Neurosurgical Review, 2020, 43(5): 1235-1253.

[2] AMIN J, SHARIF M, YASMIN M, et al. Use of Machine Intelligence to Conduct Analysis of Human Brain Data for Detection of Abnormalities in Its Cognitive Functions [J]. Multimedia Tools and Applications, 2020, 79(15/16): 10955-10973.

[3] MISHAKOVA A, PORTET F, DESOT T, et al. Learning Natural Language Understanding Systems from Unaligned Labels for Voice Command in Smart Homes [C]//2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops). Kyoto, Japan: IEEE, 2019: 832-837.

[4] 高灵霞，孙凤兰. 终端用户编程的自然语言语义解析方法研究 [J]. 西南师范大学学报(自然科学版)，2020，45(5)：134-140.

[5] YOUNG T, HAZARIKA D, PORIA S, et al. Recent Trends in Deep Learning Based Natural Language Processing Review Article [J]. IEEE Computational Intelligence Magazine, 2018, 13(3): 55-75.

[6] DONG L, YANG N, WANG W H, et al. Unified Language Model Pre-training for Natural Language Understanding and Generation [C] //33rd Conference on Neural Information Processing Systems (NeurIPS 2019). Vancouver, Canada. 2019: 1-14.

[7] 龚　静，李英杰，黄欣阳. 基于统计词典和特征加强的多语言文本分类 [J]. 西南师范大学学报(自然科学版)，2018，

43(9): 45-50.

[8] TRIPATHI S, SINGH C, KUMAR A, et al. Bidirectional Transformer Based Multi-Task Learning for Natural Language Understanding [C]//24th International Conference on Applications of Natural Language to Information Systems, NLDB 2019. Salford, UK: Springer , 2019: 54-65.

[9] WIGNELL P, CHAI K, TAN S, et al. Natural Language Understanding and Multimodal Discourse Analysis for Interpreting Extremist Communications and the Re-Use of these Materials Online [J]. Terrorism and Political Violence, 2018: 1-26.

[10] PATKI S, DANIELE A F, WALTER M R, et al. Inferring Compact Representations for Efficient Natural Language Understanding of Robot Instructions [C]//2019 International Conference on Robotics and Automation (ICRA). Montreal, QC, Canada: IEEE, 2019: 6926-6933.

[11] QIU Z M, CHO E, MA X C, et al. Graph-Based Semi-Supervised Learning for Natural Language Understanding [C]// Proceedings of the Thirteenth Workshop on Graph-Based Methods for Natural Language Processing (TextGraphs-13). Hong Kong, China. Stroudsburg, PA, USA: Association for Computational Linguistics, 2019: 151-158.

[12] LIU X D, HE P C, CHEN W Z, et al. Multi-Task Deep Neural Networks for Natural Language Understanding [C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy. Stroudsburg, PA, USA: Association for Computational Linguistics, 2019: 4487-4496.

[13] SUN F, LIU J, WU J, et al. BERT4Rec: Sequential Recommendation with Bidirectional Encoder Representations from Transformer [C]//Proceedings of the 28th ACM International Conference on Information and Knowledge Management. Beijing China: ACM, 2019: 1441-1450.

[14] ZHAO L, FENG Z. Improving Slot Filling in Spoken Language Understanding with Joint Pointer and Attention [C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Melbourne, Australia: Association for Computational Linguistics, 2018: 426-431.

Application of Distributed Semantic Framework in Natural Language Understanding

LI Xiao-wen[1], ZHU Qi-liang[2]

1. *School of Computer and Information Engineering, Shanxi Technology and Business College, Taiyuan 030006, China;*
2. *School of Information Engineering, North China University of Water Resources and Electric Power, Zhengzhou 450045, China*

Abstract: In order to learn the embedded semantic correspondence between unstructured text and its corresponding structured semantic knowledge, a distributed semantic vector learning framework has been proposed in this paper for natural language understanding (NLU). The semantic framework, with long-term memory, aims at encoding the input sequence to generate the text vector, and then at combining the intention tag, timeslot tag and timeslot value vector to generate the distributed semantic vector. By minimizing the distance between the text output vector and the semantic framework vector, the semantic equivalency vector is placed in the vector space, and finally uses the intention reconstruction and timeslot tag generation loss as the goal score is to learn the robust semantic vector. Experimental results show that the learned semantic vector contains semantic information, and the proposed semantic framework is better than the NLU system in terms of NLU results rearrangement.

Key words: natural language understanding; distributed representation; semantic vector learning; semantic framework reconstruction

责任编辑 崔玉洁