

层次概念的分布式表示和学习方法综述

朱晓光

(徐州工程学院 管理工程学院, 江苏 徐州 221018)

摘要:层次概念能够有效解释语言模型的隐含知识,并且提升语言模型的结构化优化性能。针对层次概念表示和学习模型的多样化发展,该文分析了层次概念表示的线性空间、概率空间和文本蕴含性质,梳理了概念学习模型的层次结构和优化原理,用于促进概念学习模型的应用效果。通过阐述概念学习过程和语义空间的层次性质,归纳层次概念学习的四类计算模型:基于条件熵的文本层次概念抽取;建立语言资源的概念层次和神经网络的隐含层次之间的映射;通过迭代的随机过程拓展主题模型;在正则化因子中添加语义关系约束。通过概念学习综述得出如下结论:层次性的语言模型广泛结合了显明和隐含的概念表示方法;统计模型和语言资源的语义映射是拓展层次结构的主要路径;层次结构具有双曲空间和嵌套球形结构;层次结构分析可以提升统计模型的解释水平。

关键词:层次概念;概念学习;分布式表示;统计语言模型;层次主题模型

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2023)10-0001-07

doi:10.3969/j.issn.1673-629X.2023.10.001

Survey of Distributed Representation and Learning Methods on Hierarchical Concept

ZHU Xiao-guang

(School of Management Engineering, Xuzhou University of Technology, Xuzhou 221018, China)

Abstract: Hierarchical concepts can effectively explain the implicit knowledge and improve the structural optimization performance of language models. Concerning the diversified development of hierarchical concept representation and learning models, we analyze the linear space, probability space and textual entailment properties of hierarchical concept representation, and sort out the hierarchical structure and optimization principles of concept learning models to promote the application of concept learning models. By expounding the hierarchical nature of the concept learning process and semantic space, four types of computational models for hierarchical concept learning are summarized, including text-level concept extraction based on conditional entropy, establishing a mapping between the concept level of language resources and the hidden level of the neural network, expanding the topic model through an iterative stochastic process, and adding semantic relationship constraints in the regularizer. Through the survey of concept learning, the following conclusions are drawn. Hierarchical language models widely combined explicit and implicit concept representation methods. Semantic mapping between statistical model and language resources are the main paths to expand hierarchical structures. Hierarchical structures are distributed with hyperbolic space and nested spherical structure. Hierarchical structure analysis can improve the interpretation level of statistical models.

Key words: hierarchical concept; concept learning; distributed representation; statistical language model; hierarchical topic model

0 引言

概念表示方法是语义分析的基础问题,也是影响语言模型的结构和性能的重要因素。随着计算技术的发展,层次概念和分布式表示逐渐成为了语言模型的基础组件,并且促进了显明语义和隐含语义的相互映射^[1-2]。通过结合计算语言理论和深度学习模型,各类层次概念表示方法显著提升了语言处理任务的性

能,并且有助于解释和提取统计模型的语言属性^[3]。相对而言,统计语言模型的层次概念结构还缺少系统性的分析,不能有效地促进自然语言处理模型的应用和拓展。因此,基于语义学理论和层次概念表示方法,该文综述了概念学习模型的表示方法和学习原理,解析了层次语义空间的计算原理和统计性质,讨论了层次概念计算模型的应用场景和研究建议。

收稿日期:2022-10-14

修回日期:2023-02-16

基金项目:国家自然科学基金(71774101,61802230);江苏省现代教育技术研究课题(2021-R-87014)

作者简介:朱晓光(1989-),男,讲师,研究方向为知识表示、自然语言理解。

1 语义空间的层次类型

按照层次结构的性质,语义空间可以划分为三种类型:线性空间、概率空间和文本蕴含。其中,概念的涵摄关系(Subsumption)具有不同的表示方法和分布规律。通过映射至共同的语义空间,语言资源和统计模型之间形成了相互的信息增益。

1.1 线性空间

线性空间主要刻画语义分布的几何性质。其中,分布式的概念对应于整个网络的一种稳定的激活模式。在连续的语义空间之中,若是抽象的语义场(上位词)覆盖特定的区域,则会形成层次的语义空间,如图1所示。

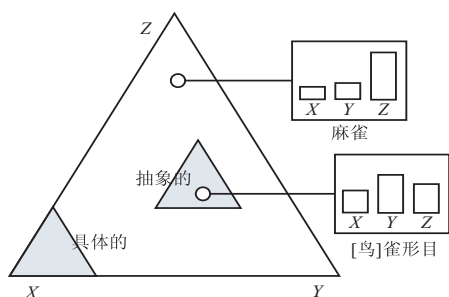


图1 上位词的区域中心性质

概念学习和层次聚类之中,上层节点包含较多的特征,因而会趋向于语义空间的中心区域,具有更为均匀的特征关联。这种区域关系形成了语义空间的指向性特征。在GloVe的预训练空间和实体关系识别模型之中,语言属性的向量都具有规律性的方向。

除了指向性特征,概念的分布区域也存在不同的形状。决策树的节点具有矩形或超矩形的概念分布。层次结构的分布式表示通常建立球形的概念分布区域^[1]。非线性的监督学习模型可能生成复杂的分类器边界。若是采用核心特征作为维度,则特征间时有互斥;若是采用样例作为维度,则概念区域一般都是凹平面。

1.2 概率空间

概念表示的概率空间主要包括各种层次性的主题模型。通过假定一组概率分布,相关模型构建了由主题生成词汇的随机过程。在递归的生成过程中,词汇的内涵逐渐趋于具体的特征分布。例如,图2生成了10个样本点,其中 p_{i+1} 是狄利克雷分布 $\text{Dir}(a = p_i)$ 所生成的一个随机样本。随机样本的相对位置表明:随机过程趋向于概率单纯形的边缘位置,获得相对具体和单一的特征。

层次主题模型体现了概念的原型化现象。对于概念分布的概率空间,词汇向量的持续归一化表明了原型的激活过程。原型的激活减少了文本的语义,从而保持概率性的词汇分布。这种现象在文本蕴含中更为

明显:词汇的累加本应增加语义,却因原型化而减少了长句的内涵。

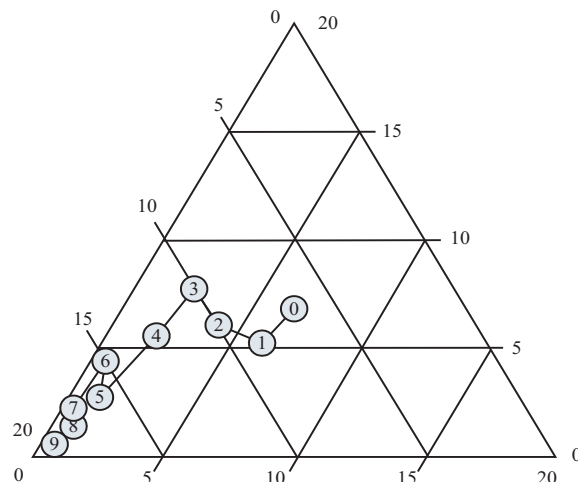


图2 狄利克雷的随机样本序列

1.3 文本蕴含关系

文本蕴含(Textual entailment)是文本之间的一种指向性的推理关系。其中,作为前提(Premise)的文本可以推出(蕴含)作为结论(Conclusion)的文本。整句蕴含主干句,则形成言语表达的层次性和构成句法。添加补充语或细化词汇可以形成句子的层次关系。补充语或下位词将限定句子的所指,由此蕴含原始的句子。图3给出了一个词汇细化的例子。其中,句子“保持指标在合理区间”的“指标”可以逐渐细化。当然,也可以按照句法添加补语,生成涵义更为具体的句子。

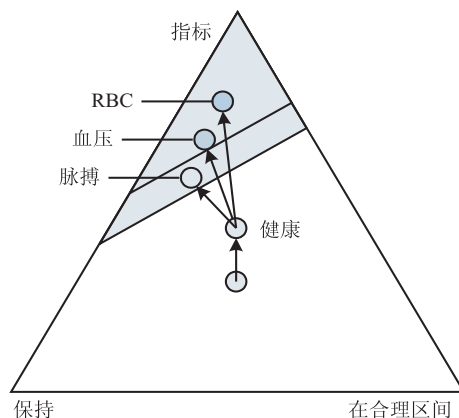


图3 逐步细化的句子内涵

补充语的添加需要遵守句法,否则,将不能直接约束句子的内涵,继而形成文本蕴含的关系。换言之,只有按照句法添加修饰语才可以细化句子的内涵。其中,句法主要指构成句法(Constituency syntax)。词汇的组合可以视为一个单元或成分(Constituent),并且逐层地向上组合语义。

2 层次概念的计算模型

参照语义组约束和概念相对化的理论,本节综

述和归纳层次概念的表示方法和学习模型,主要包括:显明的文本层次结构、神经网络模型、层次概率主题模型和语言资源正则约束。

2.1 显明的文本层次结构

文本的语言单位之间存在显明的层次关系。这种关系主要通过统计特征和语言学模板进行提取,其结果是词汇、词组和知识库条目组成的层次概念结构。这种方式仅仅使用显明的语义表示,因此更加容易理解和表达,不会出现各类优化和拟合的问题。但是,对于开放的规模性文本,显明构造具有高成本和低召回率的问题。

2.1.1 词汇共现

根据词汇的共现和句法特征,相关研究构建了文本的层次结构发现方法^[4]。基于词汇的 ID、TF-IDF、PMI 的属性组合,可以判断“网络”是“社交网络、金融网络”的上位词汇,由此拓展层次关系的语言资源^[5]。此类“主词蕴含复合词”的层次关系主要通过度量共现频率获得,例如:给定阈值 t ,如果 $P(x|y) \geq t, P(y|x) < t$,则称概念 x 部分蕴含概念 y ^[6-7]。其中,上位词一般标记为主词或抽象类别,例如,抽象类别涵盖相对具体的类别:“Mammal→Cat, Whale”^[8];概称词汇或主词作为上层节点涵盖复合词汇:“market→price→stock price”。

2.1.2 语言学模板

词汇的语言属性和知识类型也可以辅助层次结构识别。层次关系构建应用了语言学模板、依存关系类别及搭配实例^[9-10],并且通过度量概念范围和概念距离指数等概念以规范和指导层次结构的资源构建^[11]。句式可以指示词汇之间的上下位关系,例如:“A 包括 B”“B 属于 A”。通过这些典型句式,可以发现“北京/城市、石油/资源”等层次关系^[12]。此外,通过度量专业词汇的领域相关性、内聚性和结构关联,金融、生物等领域的知识类型和词法也可以提取层次关系^[13]。

如果使用语言资源作为层次结构的模板,那么显明文本结构分析还可以拓展语言资源,建立语言资源和文本数据之间的语义映射。语言资源包括通用词典、领域知识库和百科全书,其中包含了义项、词条和领域对象的层次关系。基于知识库的条目分布,显明语义分析(ESA)可以拓展知识库的语义关系^[14],例如:推特数据扩展的词条层次结构“Economics → Economics organizations → wiki; en: World Econ. Forum”属于推特分类法的一条路径^[15]。其中,词汇的多重表示可以连接 WordNet、Freebase、维基百科等语言资源中的层次关系。

2.2 神经网络模型

神经网络具有多层次的隐含特征空间,根据隐含

层的节点数量差异形成递进的抽象层次。由于存在连续性和深度结构的特征,通常需要人机结合的方式提取神经网络模型的隐含知识和层次概念^[16]。此类研究包含对偶的两个部分。一是根据已有的分类法和句法构建神经网络模型,提升自然语言处理的性能。二是根据优化的模型参数提取新的、概率性的语言属性。

2.2.1 分布式语义空间

神经网络模型的层次概念主要通过点、模、区域和方向进行表示和运算。词汇可以嵌入线性空间,通过实数向量表示各类语言单位。基于词嵌入表示,神经语言模型(NLM)训练了文本语料的分布式语义空间,其中,层次结构的节点之间存在逐层次的指向性特征。相对于下位词,上位词可以逐渐趋于原点,由此形成区域之间的层次性质,例如:通过线性变换 $y = \Phi x$,下位词 x 映射至其上位词 y ^[17]。

层次概念的区域表示主要采用嵌套球形和双曲空间结构^[18-19]。在基于区域关系的 N-Balls 模型之中,一个 N-ball 可以表示为: $w = B(\vec{A}_w, r_w)$,即:通过中心点 \vec{A}_w 和半径 r_w 表示一个词义 w 。N-ball 之间具有包含、分离两种关系和余弦距离度量方法^[1]。双曲空间采用反双曲余弦函数作为概念相似的度量方法,整个语义空间呈现边缘扩散的形态^[20]。双曲空间类似于主题生成模型,高纯度(Purity)的区域将会作为低纯度区域的子类^[21]。

2.2.2 层次性输出结构

神经网络模型的输出层可以应用显明的树形结构,称为分类树或层次柔性函数(Softmax function)。神经语言模型可以按照层次结构的路径逐层进行词汇或类别的预测。结合 WordNet、聚类算法和非线性约简方法,层次神经语言模型和对数线性模型较早地构建了层次性的输出结构^[22-23]。其中,目标词汇预测采用了词汇所在路径的概率,即:路径中所有节点的概率乘积。节点的概率主要采用相对于兄弟节点的柔性函数。

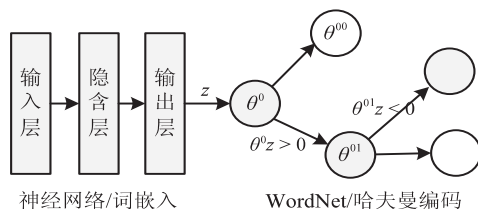


图 4 神经网络的层次性输出结构

层次性输出多数采用了二分树结构,如图 4 所示。其中,层次输出的一个隐含节点代表一个分类器,按照 $\theta^i z > 0$ 判断子节点激活^[24-25]。层次输出的每个节点蕴含一个分类边界,可以表示目标概念的区域性分布特征。

层次输出结构具有多种来源,主要包括语言资源、标签群组和多层次文本分类^[26]。层次性语言资源可以直接提供标签预测的输出结构,用于逐层地判断标签所属的节点^[27]。相对于神经语言模型,文本分类通常应用多子节点的层次结构,例如:标签群组的子集关系、多分枝(K-ary)的树形结构的贝叶斯层次聚类算法^[28-29]。相对于序列化的语义解码,树形结构的解码器能够层次化地捕获组合语义,并且保证解码过程中不同组成部分的独立性,进而更好地生成语义框架^[30]。

层次的输出层结构可以在优化过程中进行动态调整。基于信念网络的层次隐含类别(HLC)模型首先通过层次聚类获得初步的结构,然后根据 BIC 调整两侧的节点数量^[31]。通过增加剪枝和校正的过程,分布式的聚类树也可以动态适应新词的训练^[25,32]。

2.2.3 拓展的神经网络结构

神经网络模型也被广泛用于语言模型的训练和层次结构的优化。针对语言模型,相关研究主要进行了如下拓展:(1)根据情境理论应用注意力机制;(2)按照句法结构递归地进行张量运算;(3)按照层次结构建立神经网络模型,并且在各层次的模型之间共享参数。这些拓展结构虽然存在优化方法上的差异,但是概念表示主要还是应用了词汇、位置、语言属性的分布式表示方法。

神经网络的集成模型和显明层次结构之间可以逐层建立语义映射,如图 5 所示。对于层次标签预测任务,HFT-CNN 针对每一层标签训练一个卷积神经网络模型,并且在模型之间共享卷积层的参数^[33]。基础层次映射是语言单位的粒度层次,主要针对词汇、句子和语篇等语言单位分别建立卷积网络,根据语言单位之间的包含关系约束模型之间的共享参数^[34]。在训练过程中,各层节点的组合约束使得概念不仅要靠近自身类别的原型,同时也要逼近其祖先类别^[35]。

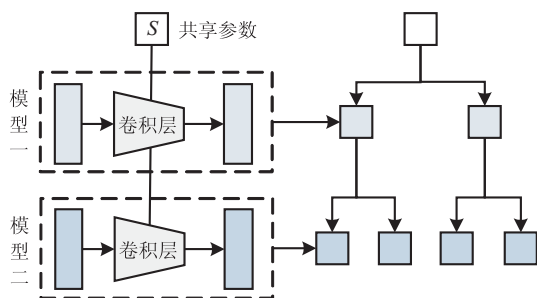


图 5 组合模型和概念层次之间的映射

循环神经网络和卷积神经网络主要引入了注意力模型、上下文权重和多粒度语言单位。其中,卷积单元和 LSTM 单元都可以作为层次语言模型的节点^[36]。通过应用局部学习规则(Oja's Rule),前馈的层次脉

冲网络也被用于学习概念的层次结构^[37]。根据语言单位的粒度性质,层次注意力网络按照词汇和句子的包含关系构建了神经网络模型,并且逐层添加注意力单元^[38]。基于句子的构成(Constituency)句法,递归神经网络可以灵活地按照节点组合语义,完成情感识别和句法解析等任务^[39-40]。针对语言模型解释的需要,R2D2 融合了组合句法和分布式表示,通过递归的转换器解析了句子的构成句法^[39]。更加复杂的神经网络则是性能导向的,可能具有层次的模型形态,但是并不蕴含概念的层次结构^[41]。

2.3 层次概率主题模型

层次概念的概率模型主要是指层次主题模型(HTM)及其拓展模型。词汇频率符合多项式分布模型,因此,主题模型采用了词汇的先验分布狄利克雷分布,例如:狄利克雷过程的向上合并构建概率混合模型和嵌套的狄利克雷过程^[42-43]。对于无监督的概念学习过程,嵌套的中餐馆随机过程(nCRP)经常被用于计算深度的、具有动态结构的主题树^[44]。在计算路径概率的过程中,nCRP 可以动态适应主题范围的变化^[45]。与此类似,BRT 模型的节点可以展开任意数量的分支,其中,父节点的概率分布部分地约束子节点的分布^[46]。贝叶斯树(BRT)也被用于分类法的构建,其中,层次树的非叶子节点表示词汇的划分,包含词汇和上下文的集合^[47]。

类似于图 2 所示的生成过程,层次性随机过程的样本将逐渐偏向具体概念。在主题的概率空间之中,主题树的根节点是相对均匀分布的,叶子节点则指示具体的主题^[48-49]。文本主题空间的层次结构之中,第一层预测元主题群组,并且在群组内进一步细化主题的分类^[50]。层次概率模型可以结合不同种类的语言单位和语言属性,形成多粒度的语言模型。除了基础的词汇分布,句子、篇章等语言单位都可以产生自一个随机过程。在词汇主题(Topic)的基础上还可以添加句子的主题(Theme)^[51]。层次隐含树分析(HLTA)模型之中,叶子节点表示“(可观察的)词汇是否出现在文档之中”,非叶子节点表示“(未观察到的)文档属性”^[49]。

概率性层次概念的主要优势在于显明的随机过程和组合方式。概率性的层次概念可以相互结合,用于提升层次概念学习的性能。通过组合词汇的概率分布和语义标签提升主题的内聚性和稳定性^[52]。HMTC 通过层次关系的语义标签和分布式表示提升了层次文本分类的性能^[26]。

2.4 语言资源的正则约束

语言资源的层次结构可以拓展概念表示学习的成本函数。通过添加节点之间的蕴含关系和基于路径的

语义相似度,拓展的成本函数可以限定相似的或互斥的概念在语义空间中的间距。语言单位的粒度具有基础的语义约束,例如:句子与其成分的语义相似。概念学习的基础的语言单位是词汇,其次是具有不同粒度的语言单位,主要包括:字素、字符、词素、句子和篇章^[53-55]。不同粒度的语言单位自然地形成层次结构,可以作为神经网络模型的内部结构和约束条件,用于提升概念学习的性能。

成本函数的拓展意味着语义分布受到多重的约束,如图6所示。在构建分布式语义空间的过程中,文本共现作为主要的远程监督,语言资源约束则是成本函数的扩展,例如,百科全书约束实体的分布式层次关系^[56]。相反,在扩展语言资源的过程中,文本共现作为辅助部分来提升显明语义的特征完备性,例如,知识表示综合的考虑实体属性、层次类型和文本关系^[57]。

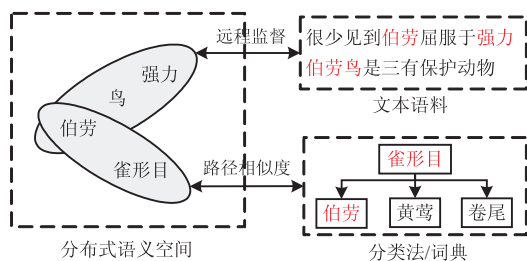


图6 概念学习的多重语义约束

除了言语表达的语言单位,语言资源也包含了特异性的语言单位和语义属性,例如:WordNet的义项(Synset)、HowNet的义素(Sememe)和概念基元^[58-59]。这些语言单位的层次结构可以直接用于分布式表示的优化,通过语义相似度的约束使得相邻的节点具有相似的语义分布,即:层次结构关系限定语义空间之中的词汇间距^[60]。由此,分类法的父节点和子节点处于连续语义空间中的邻近区域^[61]。根据预先设置的图像标签的层次结构,层次特征选择模型(HFSGR)添加了 $L_{2,1}$ -norm正则因子,使得相近的子树具有近似的稀疏表示^[62]。除了词汇间距的约束项,预测节点和实际节点在分类法中的路径距离也可以作为成本函数的约束项^[19]。

3 结束语

根据计算语言模型,该文综述了层次概念的分布式表示特征和学习模型,归纳了语义空间的层次结构和概念学习类型。对于层次概念的学习,显明文本结构分析容易提取规范文本的层次结构,但是不包含深层语义。神经网络的层次模型主要依赖概念的分布式表示,通过深层语义和语言资源的映射显著地提升模型性能。神经网络的层次关系不具备直接解释的途径,必须参照语言资源进行知识提取。层次主题模型

采用嵌套的随机过程,其层次概念具有概率分布的性质和解释方法。

层次概念学习模型正在朝三个方面发展。第一,通过建立语言资源和分布式表示之间的语义映射,提升语义空间的稳定性和完备性。第二,提取计算模型中的层次结构和分布性质。针对知识提取和转移学习的需要,逐步解释统计模型的语言属性和知识类型,促进中文层次概念的构造和语言资源的拓展。第三,建立计算模型和语言理论之间的映射,进一步完善原型、语义框架和层次关系的分布式表示方法,解释对应的几何运算和语义组合现象,建立各类语义空间和语言单位之间的内在联系,促进语言资源的融合发展。

参考文献:

- [1] DONG T, BAUCKHAGE C, LI J, et al. Imposing category trees onto word-embeddings using a geometric construction [C]//International conference on learning representations. New Orleans: ICLR, 2019: 1-10.
- [2] BENGIO Y, COURVILLE A, VINCENT P. Representation learning: a review and new perspectives [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35 (8): 1798-1828.
- [3] 汪诚愚, 何晓丰, 宫学庆, 等. 面向上下位关系预测的词嵌入投影模型 [J]. 计算机学报, 2020, 43 (5): 868-883.
- [4] SANDERSON M, CROFT B. Deriving concept hierarchies from text [C]//Proceedings of the 22nd annual international ACM SIGIR conference on research and development in information retrieval. New York: ACM, 1999: 206-213.
- [5] URENA P. Learning IS-A relations from specialized-domain texts with co-occurrence measures [J]. Journal of Computer-Assisted Linguistic Research, 2018, 2 (1): 21-38.
- [6] DE KNIJFF J, FRASINCAR F, HOGENBOOM F. Domain taxonomy learning from text: the subsumption method versus hierarchical clustering [J]. Data & Knowledge Engineering, 2013, 83: 54-69.
- [7] MEIJER K, FRASINCAR F, HOGENBOOM F. A semantic approach for extracting domain taxonomies from text [J]. Decision Support Systems, 2014, 62: 78-93.
- [8] WANG L, WANG Y, XIE Y. Learning word hierarchical representations with neural networks for document modeling [J]. Journal of Experimental & Theoretical Artificial Intelligence, 2020, 32 (3): 515-531.
- [9] 唐琳, 郭崇慧, 陈静锋, 等. 基于中文学术文献的领域本体概念层次关系抽取研究 [J]. 情报学报, 2020, 39 (4): 387-398.
- [10] 李娟. 汉语近义词辨析知识库构建研究 [J]. 北京大学学报: 自然科学版, 2022, 58 (1): 106-112.
- [11] 熊回香, 叶佳鑫. 面向科技文献检索的关键词层次结构构建 [J]. 情报理论与实践, 2022, 45 (9): 157-163.

- [12] SNOW R, JURAFSKY D, NG A Y. Learning syntactic patterns for automatic hypernym discovery [C]//Advances in neural information processing systems. Vancouver: NIPS, 2004:1297–1304.
- [13] HOXHA J, JIANG G, WENG C. Automated learning of domain taxonomies from text using background knowledge [J]. Journal of Biomedical Informatics, 2016, 63:295–306.
- [14] GABRILOVICH E, MARKOVITCH S. Computing semantic relatedness using wikipedia-based explicit semantic analysis [C]//International joint conference on artificial intelligence. Hyderabad: IJCAI Organization, 2007:1606–1611.
- [15] FARALLI S, STILO G, VELARDI P. Automatic acquisition of a taxonomy of microblogs users' interests [J]. Journal of Web Semantics, 2017, 45:23–40.
- [16] SEVESO A, MERCORIO F, MEZZANZANICA M. A human-AI teaming approach for incremental taxonomy learning from text [C]//International joint conference on artificial intelligence. Montreal: IJCAI Organization, 2021: 4917–4918.
- [17] FU R, GUO J, QIN B, et al. Learning semantic hierarchies: a continuous vector space approach [J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2015, 23(3):461–471.
- [18] DONG T. Precise embodying tree structures onto their latent feature vectors [J]. Journal of Visual Language and Computing, 2022, 2022(1):16–24.
- [19] MANZOOR E, LI R, SHROUTY D, et al. Expanding taxonomies with implicit edge semantics [C]//Proceedings of the web conference 2020. Taipei, China: ACM, 2020: 2044–2054.
- [20] NICKEL M, KIELA D. Learning continuous hierarchies in the lorentz model of hyperbolic geometry [C]//International conference on machine learning. Long Beach: JMLR, 2018: 1109–1117.
- [21] AMALAMAN P K, EICK C F, WANG C. Supervised taxonomies – algorithms and applications [J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(9):2040–2052.
- [22] MNIH A, HINTON G E. A scalable hierarchical distributed language model [C]//Advances in neural information processing systems. Vancouver: NIPS, 2008:1–8.
- [23] BLITZER J, WEINBERGER K Q, SAUL L K, et al. Hierarchical distributed representations for statistical language modeling [C]//Advances in neural information processing systems. Montreal: NIPS, 2005:185–192.
- [24] BEYGEZIMER A, LANGFORD J, LIFSHITS Y, et al. Conditional probability tree estimation analysis and algorithms [C]//Proceedings of the 25th conference on uncertainty in artificial intelligence. Montreal: AUAI, 2009: 51–58.
- [25] XU Z, KUSNER M J, WEINBERGER K Q, et al. Cost-sensitive tree of classifiers [C]//International conference on machine learning. Atlanta: JMLR, 2013:133–141.
- [26] MA Y, LIU X, ZHAO L, et al. Hybrid embedding-based text representation for hierarchical multi-label text classification [J]. Expert Systems with Applications, 2022, 187(7):115905.
- [27] DEKEL O, KESHET J, SINGER Y. Large margin hierarchical classification [C]//International conference on machine learning. New York: ACM, 2004:27.
- [28] JERNITE Y, CHOROMANSKA A, SONTAG D. Simultaneous learning of trees and representations for extreme classification and density estimation [C]//International conference on machine learning. Sydney: PMLR, 2017:1665–1674.
- [29] BASAK J, KRISHNAPURAM R. Interpretable hierarchical clustering by constructing an unsupervised decision tree [J]. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(1):121–132.
- [30] 赵睿卓, 高金华, 孙晓茜, 等. 基于树形语义框架的神经语义解析方法 [J]. 中文信息学报, 2021, 35(1):9–16.
- [31] HARMELING S, WILLIAMS C K I. Greedy learning of binary latent trees [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(6):1087–1097.
- [32] LE H S, OPARIN I, MESSAOUDI A, et al. Large vocabulary SOUL neural network language models [C]//Annual conference of the international speech communication association. Florence: ACISC Association, 2011:1469–1472.
- [33] SHIMURA K, LI J, FUKUMOTO F. HFT-CNN: learning hierarchical category structure for multi-label short text categorization [C]//Conference on empirical methods in natural language processing. Brussels: ACL, 2018:811–816.
- [34] KALCHBRENNER N, BLUNSOM P. Recurrent convolutional neural networks for discourse compositionality [C]//Annual meeting of the association for computational linguistics. Sofia: ACL, 2013:119–126.
- [35] 任 权. 基于原型网络的细粒度实体分类方法 [J]. 中文信息学报, 2020, 34(12):65–72.
- [36] JOËL L, TOUSSAINT Y, RAÏSSI C, et al. Cross-corpus training with tree LSTM for the extraction of biomedical relationships from text [C]//International conference on learning representations. Vancouver: ICLR, 2018:1–11.
- [37] LYNCH N, MALLMANN-TRENN F. Learning hierarchically-structured concepts [J]. Neural Networks, 2021, 143:798–817.
- [38] YANG Z, YANG D, DYER C, et al. Hierarchical attention networks for document classification [C]//Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics. San Diego: ACL, 2016:1480–1489.
- [39] HU X, MI H, WEN Z, et al. R2D2: recursive transformer based on differentiable tree for interpretable hierarchical language modeling [C]//Annual meeting of the association for

- computational linguistics. Bangkok; ACL, 2021; 4897–4908.
- [40] 费 豪, 姬东鸿, 任亚峰. 基于动态句法剪枝机制的中文语义角色标注[J]. 计算机学报, 2022, 45(8): 1746–1764.
- [41] 李 豫, 周光有. 基于层次化语义框架的知识库属性映射方法[J]. 中文信息学报, 2022, 36(2): 49–57.
- [42] HELLER K A, GHAHRAMANI Z. Bayesian hierarchical clustering[C]//International conference on machine learning. New York; ACM, 2005: 297–304.
- [43] RODRÍGUEZ A, DUNSON D B, GELFAND A E. The nested Dirichlet process[J]. Journal of the American Statistical Association, 2008, 103(483): 1131–1144.
- [44] BLEI D M, GRIFFITHS T L, JORDAN M I. The nested chinese restaurant process and Bayesian nonparametric inference of topic hierarchies[J]. Journal of the ACM, 2010, 57(2): 1–30.
- [45] MA J, CUI P, WANG X, et al. Hierarchical taxonomy aware network embedding[C]//Special interest group on knowledge discovery and data mining. New York; ACM, 2018: 1920–1929.
- [46] BLUNDELL C, TEH Y W, HELLER K A. Bayesian rose trees[C]//Proceedings of the 26th conference on uncertainty in artificial intelligence. Catalina Island; AUAI, 2010: 65–72.
- [47] SONG Y, LIU S, LIU X, et al. Automatic taxonomy construction from keywords via scalable bayesian rose trees[J]. IEEE Transactions on Knowledge and Data Engineering, 2015, 27(7): 1861–1874.
- [48] CHIEN J T. Bayesian nonparametric learning for hierarchical and sparse topics[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2018, 26(2): 422–435.
- [49] CHEN P, ZHANG N L, LIU T, et al. Latent tree models for hierarchical topic detection[J]. Artificial Intelligence, 2017, 250: 105–124.
- [50] WEIGEND A S, WIENER E D, PEDERSEN J O. Exploiting hierarchy in text categorization[J]. Information Retrieval, 1999, 1(3): 193–216.
- [51] CHIEN J T. Hierarchical theme and topic modeling[J]. IEEE Transactions on Neural Networks and Learning Systems, 2016, 27(3): 565–578.
- [52] VIEGAS F, CUNHA W, GOMES C, et al. CluHTM – semantic hierarchical topic modeling based on CluWords[C]//Annual meeting of the association for computational linguistics. Seattle; ACL, 2020: 8138–8150.
- [53] SENGUPTA A, BHATTACHARJEE S K, CHAKRABORTY T, et al. HIT – a hierarchically fused deep attention network for robust code-mixed language representation[C]//Annual meeting of the association for computational linguistics. Bangkok; ACL, 2021: 4625–4639.
- [54] 李 纲, 余 辉, 毛 进. 基于多层语义相似的技术供需文本匹配模型研究[J]. 数据分析与知识发现, 2021, 5(12): 25–36.
- [55] 丁 玲, 向 阳. 基于分层次多粒度语义融合的中文事件检测[J]. 计算机科学, 2021, 48(5): 202–208.
- [56] HU Z, HUANG P, DENG Y, et al. Entity hierarchy embedding[C]//Annual meeting of the association for computational linguistics. Stroudsburg; ACL, 2015: 1292–1300.
- [57] TANG X, CHEN L, CUI J, et al. Knowledge representation learning with entity descriptions, hierarchical types, and textual relations[J]. Information Processing & Management, 2019, 56(3): 809–822.
- [58] ZENG X, YANG C, TU C, et al. Chinese LIWC lexicon expansion via hierarchical classification of word embeddings with sememe attention[C]//AAAI conference on artificial intelligence. New Orleans; AAAI, 2018: 1–8.
- [59] 吴佐衍, 王 宇. 基于 HNC 理论的词语相似度计算[J]. 中文信息学报, 2014, 28(2): 37–43.
- [60] DONG H, WANG W, HUANG K, et al. Automated social text annotation with joint multilabel attention networks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(5): 2224–2238.
- [61] YANG S H, KOLCZ A, SCHLAIKJER A, et al. Large-scale high-precision topic modeling on twitter[C]//Special interest group on knowledge discovery and data mining. New York; ACM, 2014: 1907–1916.
- [62] TUO Q, ZHAO H, HU Q. Hierarchical feature selection with subtree based graph regularization[J]. Knowledge – Based Systems, 2019, 163: 996–1008.