

复杂异构数据的表征学习综述

蹇松雷 卢凯

国防科技大学计算机学院 长沙 410073

(jiansonglei@nudt.edu.cn)



摘要 随着智能时代和大数据时代的到来,各种复杂异构数据不断涌现,成为数据驱动的人工智能方法、机器学习模型的基础。复杂异构数据的表征直接关系到后续模型的学习性能,因此如何有效地表征复杂异构数据成为机器学习的一个重要研究领域。文中首先介绍了数据表征的多种类型,并提出了现有数据表征方法面临的挑战;其次,根据数据类型将数据划分成单一类型数据和复合类型数据,针对单一类型数据,分别介绍了4种典型数据的表征学习发展现状和代表算法,包含离散数据、网络数据、文本数据和图像数据;然后,详细介绍了4种由多个单一数据或数据源复合而成的复杂数据,包含了离散特征与连续特征混合的结构化数据、属性数据与复杂网络复合的属性网络数据、来自不同领域的跨领域数据和由多种数据类型复合的多模态数据,分别介绍了基于上述复杂数据的表征学习现状以及最新的表征学习模型;最后,对复杂异构数据表征学习的发展趋势进行了探讨。

关键词: 表征学习;机器学习;离散数据;属性网络;跨领域数据;多模态数据

中图法分类号 TP181

Survey on Representation Learning of Complex Heterogeneous Data

JIAN Song-lei and LU Kai

College of Computer, National University of Defense Technology, Changsha 410073, China

Abstract With the coming of the eras of artificial intelligence and big data, various complex heterogeneous data emerge continuously, becoming the basis of data-driven artificial intelligence methods and machine learning models. The quality of data representation directly affects the performance of following learning algorithms. Therefore, it is an important research area for representing useful complex heterogeneous data for machine learning. Firstly, multiple types of data representations were introduced and the challenges of representation learning methods were proposed. Then, according to the data modality, the data were categorized into single-type data and multi-type data. For single-type data, the research development and typical representation learning algorithms for categorical data, network data, text data and image data were introduced respectively. Further, the multi-type data compounded by multiple single-type data were detailed, including the mixed data containing both categorical features and continuous features, the attributed network data containing node content and topological network, cross-domain data derived from different domains and the multimodal data containing multiple modalities. And based on these data, the research development and state-of-the-art representation learning models were introduced. Finally, the development trends on representation learning of complex heterogeneous data were discussed.

Keywords Representation learning, Machine learning, Categorical data, Attributed network, Cross-domain data, Multimodal data

1 引言

数据的迅速增长和机器智能的发展,催生了智能时代的到来,而数据无疑是一切发展和变革的基础。数据不仅驱动人工智能和机器学习的发展,也挑战着现有机器学习算法的

学习能力。传统的机器学习算法主要针对单一类型、结构化的数据,如监督学习的经典算法 SVM、决策树、无监督学习的经典算法 k-means 聚类等。而现在的大数据大多是复杂异构的,对于这种数据,一方面需要改进现有的机器学习算法,另一方面需要统一数据表征,将复杂异构数据转化为普通机器

到稿日期:2019-06-28 返修日期:2019-08-26 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家重点研发项目(2018YFB0803501);国防科技卓越人才计划(2017-JCJQ-ZQ-013);国家自然科学基金(61902405);湖南省科技领军人才计划(2017RS3045)

This work was supported by National Key Research and Development Program of China (2018YFB0803501), National High-level Personnel for Defense Technology Program (2017-JCJQ-ZQ-013), National Natural Science Foundation of China (61902405) and Hunan Province Science Foundation (2017RS3045).

通信作者:卢凯(kailu@nudt.edu.cn)

学习算法能够处理的连续型向量数据。

数据表征或数据表示(Data Representation)是机器学习与数据之间的桥梁,机器学习算法在数据表征的基础上进行优化、预测,从而拟合数据分布或学习数据表征到标签(label)的映射。在这个过程中,一个有效的数据表征对机器学习的结果至关重要,因为机器学习算法很大程度上依赖于数据表征的有效性^[1]。

通常可以将数据表征分成两个大类,一类为离散表征(Categorical Representation),另一类为连续表征(Continuous Representation),而连续表征又可以分为分布表征(Distributional Representation)和分布式表征(Distributed Representation)。离散表征指将数据用离散值表示,例如,在文本数据中用单词的ID来表示一个词;在结构化数据中用0或1表示真或假。离散表征通常是有限的,离散值本身没有大小的区分,例如可以用0和1表示男性和女性,同样也可以用10和20表示男性和女性。离散表征一般只是将数据符号化,而不包含任何语义信息。通常,在实际算法中用独热编码(one-hot encoding)来转换离散表征,即表征向量的长度等于所有可能取值的个数,每一位对应一种可能的取值,如果某位为1,则表示数据的取值为该位对应的值。经典的机器学习算法,如传统聚类、高斯混合模型、最近邻方法、决策树等,都是基于这样的表征。例如,在经典聚类算法k-means聚类中,数据对象的类别由最近的一个中心点指定,这个类别的表征就是典型的离散表征。

分布表征基于分布假设理论,利用数据的概率分布来表征数据。分布表征在文本表征中较为常见,它是基于分布假说的^[2],即上下文相似的词,其语义也相似,如潜在语义分析模型(Latent Semantic Analysis,LSA)、潜在狄利克雷分配模型(Latent Dirichlet Allocation,LDA)等。基于分布表征的学习模型一般是生成式模型,即求得观测数据变量和目标变量(通常为标签数据)的联合概率分布。通常,建模求解数据的分布表征一般基于很强的假设,且模型的求解或推断较为复杂。

分布式表征是指将数据嵌入到低维连续空间,通常称这种表征为嵌入(embedding)或向量(vector)。相比于传统的独热表征,同样长度的分布式表征能表达更多的信息。例如,独热表征需要 $O(N)$ 个参数才能区分 $O(N)$ 个输入空间,而如果采用分布式表征, $O(N)$ 个参数能够区分 $O(2^N)$ 个输入空间。目前,学习数据的分布式特征或嵌入的常用方法是人工神经网络,例如,自然语言处理中的词嵌入学习的经典方法Word2vec^[3]、复杂网络中节点嵌入的方法Node2vec^[4]等。本文所讨论的数据表征也都是分布式表征。

表征学习或表示学习(Representation Learning),指在构造分类器或者预测器时从数据中抽取有用信息,从而学习到数据表征^[1]。表征学习强调数据表征为预测服务,因此一个好的表征需要为预测器提供有用的信息。从数据的角度来说,表征需要提取数据的有效信息,能够反映数据的真实分布。从算法的角度来说,表征需要满足算法的应用格式,方便算法的学习,提高预测的准确率。

对于复杂异构数据,表征学习表现出了它的必要性和优

势。在面对复杂异构数据时,只能通过表征学习统一不同类型、不同模态、不同结构的数据,解决其中的分布异构性、结构异构性和模态异构性等问题,从而为预测算法提供有效的数据表征。相关的表征学习模型已经在迁移学习(Transfer Learning)、领域适应(Domain Adaptation)以及多任务学习(Multi-task Learning)中取得良好的效果^[5-6]。但是,现阶段的表征学习也面临诸多挑战:

(1)如何在数据表征中提取和学习结构化数据内部复杂的耦合关系。传统的分布式表征假设数据特征是独立同分布的,即假设特征为一组独立同分布的变量,而在实际数据中,数据特征之间存在大量的交互和耦合关系。

(2)如何在数据表征中综合提取和学习关系型数据和非关系型数据中的有用信息。关系型数据强调的是数据对象之间的交互,是非关系型数据(如结构化数据、图像、视频、文本等)无法替代的一类数据,且二者之间的信息相互交融,难以在表征中统一表示。

(3)如何在数据表征中提取和学习不同领域中的有用信息。不同领域的数据一般来自不同的信息源,但是它们的模态可能相同。跨领域数据面临更多的非一致信息,表征学习模型需要根据任务的假设提取一致信息或协调非一致信息,从而实现不同领域之间的相互学习和借鉴。

(4)如何在数据表征中提取和学习不同模态中的有用信息。不同模态的数据(包含结构化数据、图像、视频、语音、文本等)可能来自于相同或者不同的信息源。各个模态数据虽然可以通过各自独立的表征学习方法将数据映射到实值向量空间,但是各个向量空间包含的语义信息却各不相同。

近年来,随着深度学习的发展,表征学习领域也取得了很大的进展,本文将介绍多种典型的复杂异构数据的表征学习现状。本文第2节介绍了4种典型的单一类型数据的表征学习模型,包含离散数据、网络数据、文本数据和图像数据;第3节介绍了由这4种单一类型数据复合而成的异构数据表征学习模型,包含混合数据、属性网络数据、跨领域数据和多模态数据;第4节对未来复杂异构数据表征学习的发展方向进行了探讨。

2 单一类型数据表征学习模型

2.1 离散数据的表征学习

离散数据是典型的属性数据或结构化数据,它一般通过二维的数据表表示,每一行表示一个数据对象,每一列表示一个特征,一个数据对象在一个特征上所对应的值称为特征值。在属性数据中,根据特征的不同类型又可将数据分为连续数据、离散数据和混合数据。其中,连续数据是天然的数据表征,一般不需要进行转化,但可以对它进行降维、提取等操作。

在离散数据表征中,编码(encoding)是最普遍的一种表征方式^[7],其中最流行的方法是独热编码,又称为一位有效编码,即将每一列离散特征编码为一个二值矩阵(通常是0和1)。例如,将“性别”这个离散特征用两位的向量来表示,当特征值为“男性”时,编码为“10”,当特征值为“女性”时,编码为“01”。尽管通过独热编码可以将离散数据转化为连续数据,但是它隐含的假设是所有特征值都是独立和等价的。而在实

际数据中,各个特征和特征值并非完全独立或等价。同时,独热编码会导致表征的维度非常高,从而引起维度灾难^[8]。但是可以通过一些降维方法,如主成分分析(PCA)^[9],来降低独热表征矩阵的维度。另一个比较有名的编码方式是 IDF 编码^[10],即通过每个特征值频率倒数的 log 值来表征该特征值, IDF 能从特征值出现频率的角度捕获一些数据特点。尽管这些编码方法实现简单、效率很高,但是它们都无法捕获数据中复杂的特征值耦合关系。

通过数据对象的两两相似度矩阵来表征离散数据也是一种常见的离散数据表征方法。在学习相似度时可以融入数据中的耦合关系,如 ALGO^[11]使用特征值之间的条件概率来表述特征值耦合关系;DILCA^[12]和 DM^[13]用特征选择和特征权重来衡量特征之间的耦合关系;COS^[14]考虑了特征之间和特征内部的耦合关系;CMS^[15]在 COS 的基础上提出了数据对象之间的距离度量;CURE^[16]是一个最新的通用离散数据表征框架,它在数据表征的学习过程中捕获不同层次的特征值、特征之间的耦合关系,能根据聚类 and 异常检测的特点将框架实例化为不同的表征算法,CURE 的结构如图 1 所示。其他的嵌入方法,如文献[17]提出的方法需要引入人类标签来学习对象之间的距离,使得数据表征过度拟合监督信息,从而丢失了原本的数据特点,使其在无监督任务中表现欠佳。

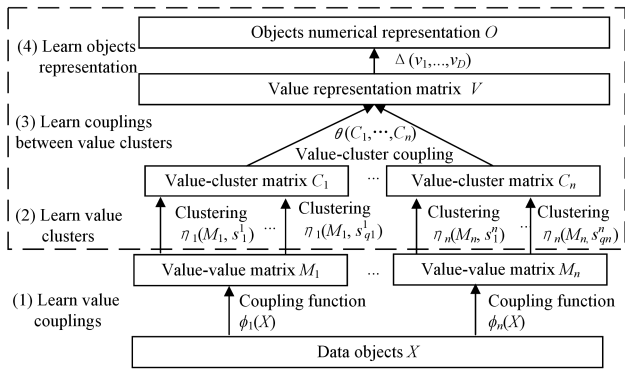


图 1 CURE 框架结构

Fig.1 Framework of CURE

2.2 网络数据的表征学习

网络数据是表征关系的一类重要数据,特别是在大数据时代,各种复杂网络以及动辄上亿的节点和边,使得传统的基于邻接矩阵或者邻接表的表征难以满足复杂的推理或预测。因此我们用低维向量来表示网络中的节点,将其作为网络的表征用于后续的预测任务,网络的这种表征叫做网络嵌入(Network Embedding)。传统的低维向量表征方法叫做图嵌入(Graph Embedding),一般采用矩阵分解实现降维,例如通过分解图的拉普拉斯矩阵或者节点的相似度矩阵来实现降维^[18-21]。

随着深度学习的兴起,现有的很多方法将深度学习模型或表征学习模型应用到网络表征中以学习节点的表征。DeepWalk^[22]是第一个基于深度学习技术的网络表征方法,它将随机游走和词嵌入(Word Embedding)相结合,通过将节点视为单词,将随机游走路径作为句子来模拟文本嵌入的场景,从而使用现有的文本嵌入模型(如 Skip-gram 和 CBOW^[3])来

学习网络嵌入。在 DeepWalk 的启发下,出现了 LINE^[23] 和 Node2vec^[4],LINE 采用广度优先搜索策略生成上下文节点,而 Node2vec 直接扩展了 DeepWalk,采用一个有偏好的随机游走过程,将深度优先搜索和广度优先搜索相结合。除此之外,SDNE^[24]使用深度自动编码器(Autoencoder)来保留局部网络结构;DNGR^[25]采用随机冲浪策略来捕获图形结构信息,然后采用去噪自动编码器学习节点嵌入。更丰富的网络表征原则、方法可以参见文献[26-27]。

2.3 文本数据的表征学习

除了独热编码,文本数据中最早采用分布表征,包含根据分布假说提出的潜在语义索引模型(LSI)^[28]、潜在狄利克雷分配模型(LDA)^[29]以及它们的一些变种方法^[30-31]。近年来,随着深度学习的兴起,基于文本的分布式表征方法,即词嵌入表征模型相继被提出。文本的分布式表征不仅能够大大降低表征维度,而且能够包含文字的语义信息,即可以通过词向量之间的距离衡量两个词之间的相似度。

最早的词嵌入表征方法可以追溯到 2003 年的工作^[32],而最近的工作 word2vec^[3]让词嵌入逐渐成为自然语言处理的标准表征。Word2vec 包含了两大模型,即 CBOW 和 Skip-gram,其中 CBOW 根据中心词周围的词来预测中心词,而 Skip-gram 则根据中心词来预测周围词。

与 word2vec 仅仅利用局部上下文窗口的方式不同,GloVe^[33]将词和词的共现次数直接编码到词语上下文的共现矩阵中,通过一个特定的带权最小二乘模型来实现分解,从而得到了高效的词向量表征方法。但上述方法都不能正确表征多义词,所以 ELMO^[34]提出根据上下文学习文本的表征,从而解决多义词的表征问题。而最近的方法 BERT^[35]基于 Transformer^[36]和两阶段训练机制(即预训练和后续调优训练),能够很好地利用文本中的上下文。最新的模型 XL-Net^[37]改进了 BERT,具体来说它采用掩码(Mark)标记的训练模式,并采用双流自注意力模型同时考虑了文本的上下文,因而在自然语言处理的多个任务中都取得了最优效果。

2.4 图像数据的表征学习

在图像的特征模型中,卷积神经网络(Convolutional Neural Network, CNN)已经成为图像的经典表征方法,替代了早期的人工特征提取方法。最早的 CNN 是 LeCun 使用误差梯度回传方法设计并训练的(该模型称为 LeNet^[38]),并在手写数字识别中取得了良好的识别效果。LeNet 也成为了 CNN 的基本结构,即由输入层、卷积层(convolutional layers)、池化层(pooling layers)、全连接层(fully connected layers)和输出层组成。

2012 年,Krizhevsky 等^[39]使用扩展了深度的 CNN(称为 AlexNet),在 ImageNet 大规模视觉识别挑战竞赛(ImageNet Large Scale Visual Recognition Challenge, LSVRC)中取得了当时最佳的分类效果,使得 CNN 在图像表征和应用领域占据主导地位。2014 年,Simonyan 等^[40]在其发表的文章中探讨了“深度”对于 CNN 网络的重要性,并提出了 VGG 模型。VGG 模型在 LSVRC-14 竞赛中得到了图像分类“指定数据”组的第 2 名,该模型证明了深度在视觉表示中的重要性。2014 年,Szegedy 等^[41]大大增加了 CNN 的深度,提出了一个

超过 20 层的 CNN 结构,称为 GoogleNet。在 GoogleNet 的结构中采用了 3 种类型的卷积操作,它的参数仅为 AlexNet 的 1/12,而且 GoogleNet 的准确率更高,在 LSVRC-14 中获得了图像分类“指定数据”组的第 1 名。2015 年和 2016 年,GoogleNet 的发明团队相继发布了新版本的 GoogleNet,称为 Inception V3^[42] 和 Inception V4^[43]。2015 年,He 等^[44]采用残差网络(Residual Networks, ResNet)来解决梯度消失的问题。ResNet 的主要特点是跨层连接,它通过引入捷径连接技术(Shortcut Connections)将输入跨层传递并与卷积的结果相加。在 ResNet 中只有一个池化层,它连接在最后一个卷积层后面。ResNet 使得底层的网络能够得到充分训练,准确率也随着深度的加深而显著提升。深度为 152 层的 ResNet 在 LSVRC-15 的图像分类比赛中获得了第 1 名的成绩。更多关于 CNN 的发展脉络和应用可以参考文献[45-46]。

3 多源复合类型数据表征学习模型

上述单一类型数据或多数据源复合可得到多类型的复杂异构数据;不同类型的属性数据复合可形成混合数据(即包含离散特征和连续特征的结构化数据);网络数据和其他非关系型数据复合(包括属性、文本、图像)可形成属性网络数据;来自不同领域的文本与文本、图像与图像可形成跨领域数据;属性-图像、属性-文本和文本-图像复合为多模态数据。因此,本文将主要介绍以下几类复合类型数据的表征学习现状,即混合数据、属性网络数据、跨领域数据以及多模态数据。

3.1 混合数据的表征学习

现有的大部分方法是通过数据转换来构建混合数据的表征。连续特征离散化是最经典的数据转换方法,例如 spectralCAT^[47]用一种自动化的方式离散化连续特征,主要通过对连续特征聚类和类标签来构造新的离散特征。但是这种转换之后的数据被当作独立的特征而应用到聚类模型中,从而忽视了不同类型数据特征之间的关系^[48-49]。Coupled-MC^[14]是最新的混合数据表征学习方法,它采用与 spectral-CAT 相同的连续数据离散化方案。不同的是, coupledMC 将连续特征转化到离散空间之后,采用离散变量的相似度来表示离散特征,然后计算各个特征之间的皮尔逊相关系数(Pearson Correlation Coefficient)。因为离散化连续变量带来了信息丢失,所以仅仅采用皮尔逊相关系数不能很好地捕获连续特征和离散特征之间的关系。除了离散化连续特征之外,还有一些其他的方法可以将离散特征转换成连续特征,如独热编码、CDE^[15]、UFT^[50]等,但是它们都不能处理混合数据中的特征异构性。

其他一些方法直接计算混合数据对象之间的距离或者相似度。K-prototype^[51]是 k-means 聚类算法的扩展,它将连续特征的欧氏距离和离散特征的汉明距离进行加权求和,从而计算混合数据之间的距离。按照 k-prototype 的模式,其他一些方法^[11,52-54]通过改变连续特征的距离计算方法或离散特征的距离计算方法,来产生针对混合数据的不同距离度量方法。它们中的大部分都是通过衡量离散化的连续特征与离散特征之间的共现关系来量化连续特征和离散特征的交互。

自动编码器已经在图像和文本数据上展示出了表征学习

能力^[55-56],它能够产生语义丰富的和可区分性的表示。但是,自动编码器主要捕获特征级别的耦合关系,因为它通过最小化重建错误来训练模型,而不能显式地增强对象之间的区分性,同时也不能直接处理混合数据。

距离度量学习是学习对象区分信息的一个重要方法,通过距离度量将训练数据集从原始空间转换到度量空间,同时保持其原有的距离关系^[57]。但是大部分的度量学习方法需要类别标签来指导训练过程,且大量的无监督度量学习方法^[58-59]不能直接应用于混合数据。最新的混合数据表征学习模型 MAI^[60]为混合数据构造了两个互补的编码空间,通过两个编码空间的相互学习实现了无监督的度量学习,从而学习到了区分性更强的混合数据表征,MAI 的模型框架如图 2 所示。MAI 中的两个编码空间分别基于不同的假设,一个假设是特征之间相互独立,另一个假设是异构特征之间紧耦合,这两个特征空间的编码相互学习最后达成统一和平衡,从而捕获到混合数据中的异构耦合关系。

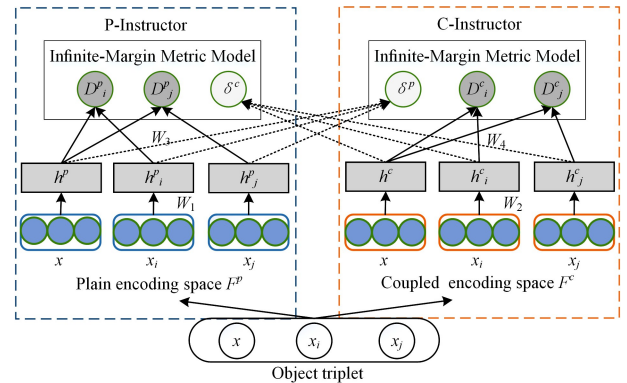


图 2 MAI 模型结构

Fig. 2 Model structure of MAI

3.2 属性网络的表征学习

属性网络指带有非关系型数据的复杂网络,如节点特征、节点标签、节点文本、节点图像或边的特征、类型等。这种关系型和非关系型共存的数据是一种典型的复杂异构数据,在实际应用中很常见。近年来,有很多工作研究如何将这种数据嵌入到低维表征中,但是属性网络中的非关系型信息、网络拓扑结构信息的深度耦合性和不同类型信息的异构性使得属性网络的表征学习面临很大的挑战。

文献[61-63]将节点标签信息引入网络表征中,通过优化分类器来优化网络表征的目标函数,从而使得同一类标签的节点比不同类标签的节点更相似。对于节点的特征信息,文献[64]将节点特征当作离散值来处理,而文献[65]则用连续高维向量来表示节点特征。文献[66]可以处理离散和连续特征混合的节点属性信息。典型的方法如 UPP-SNE^[67]假设用户属性信息和网络结构有高度相关性,因此每个节点使用它的属性特征向量来表示,通过将用户属性信息和网络结构映射到同一个表征空间实现二者的结合。

在节点属性中,另一种常见的信息是文本信息。文本表征学习的发展,如词嵌入模型 CBOW 和 Skip-gram,使得文本信息能够转换成连续向量表征,极大地方便了后续的信息结合过程。TADW^[68]模型首次通过矩阵分解将节点的文本信

息融入到网络表征中。随着网络嵌入的发展,大部分工作对现有的网络嵌入模型进行了扩展,例如文献[69]通过随机游走捕获网络结构,再在 Skip-gram 模型中加入节点内容,从而学习属性网络表征。CANE^[70]通过交互的记忆力模型(Attention Model)来考虑节点上下文对节点内容的影响,从而学习到一个上下文敏感的节点表征。文献[71]通过修改自动编码器来实现结构信息和节点属性的对齐,但这种对齐方式只会保留一致性的信息从而导致信息丢失。TriDNR^[68]同时考虑了节点的文本信息和标签信息,从而使得后续的网络表征更适应节点分类任务。而最新的方法 MAI-ECS^[72]设计了一个演化耦合模型,将网络结构和节点文本信息在网络表征中深度融合,捕获了节点之间的多方面交互关系和节点影响力,其结构如图 3 所示。MAI-ECS 不仅能产生可用于多种网络挖掘任务的网络表征,还可以学习到具有可解释性的节点影响力。

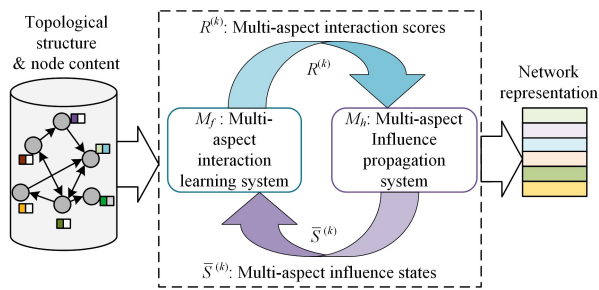


图 3 MAI-ECS:属性网络表征学习模型

Fig. 3 MAI-ECS:attributed network representation model

其他的节点属性类型也可以通过表征学习模型融入网络表征中,如图像信息^[73]、节点的位置信息^[74]以及节点用户对于产品的偏好信息^[75]等。

3.3 跨领域数据的表征学习

跨领域数据的表征学习与领域知识迁移任务紧密相关,典型的跨领域任务包括领域适应(Domain Adaptation, DA)、图像翻译以及跨领域的推荐系统。本文重点关注跨领域的图像数据以及相关的领域适应和图像翻译。跨领域数据的表征不仅需要表示出每个领域的的数据特征,还需要捕获领域之间的共性和差异性。通常,跨领域的表征学习和任务是紧耦合的,表征作为连接不同领域之间的桥梁,对知识跨领域的迁移起到了核心作用。

在跨领域数据的表征学习中,表征的解离性(Disentangle-ment)是指将数据中原本相关、耦合的因素在表征中解离成若干独立松耦合的因子^[5],能够帮助跨领域数据实现知识共享和迁移。近期的一些跨领域的表征解离工作^[76-78]大多基于对抗生成网络(GANs)^[79-80]和变分自动编码器(VAEs)^[81],通过深度学习学习可解释的表征,从而实现领域知识的迁移。文献[82]也通过单一领域的监督信号来学习跨领域的解离表征,从而构建了源领域和目标领域的共享表征空间,再利用 GAN 和分类器来综合训练神经网络。但是这样学习到的表征只包含两个领域的共同信息,而且强烈依赖于类标签。文献[83]通过跨领域的自动编码器解离两个领域的表征,构建了共享表征和各自领域独特的表征,从而可以实现领域信息的细粒度区分。文献[84]提出了联合特征解

离网络(UFDN)模型,它实现了多个领域之间的融合,构造了多领域的共享隐空间,并且解离出各个领域的独立因子,从而实现图像之间的翻译和定制化操作。文献[85]提出了新的跨领域表征学习方法,它能够从两个领域的标记数据和未标记数据中获取知识,从而为单个领域的表征获取互补信息。

3.4 多模态数据的表征学习

多模态表征学习的目标是从多种异质模态的数据(如图像-文本、图像-属性、文本-属性数据等)中提取数据对象的表征,它是多模态学习(包括模态间映射、对齐、融合和协同学习)的基础^[86-88]。多模态数据的一个典型表征方法是将各个模态的独立表征拼接在一起,形成一个联合表征,如图 4 所示,再在这个联合表征上进行后续任务的学习^[89-92]。

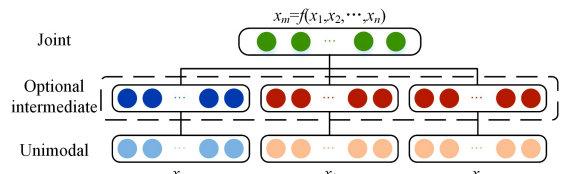


图 4 多模态数据的联合表征

Fig. 4 Joint representation of multimodal data

例如,双向深度神经网络(Bidirectional Deep Neural Nets, BiDNN)^[93]将图像和文本的表征进行联合表示。文献[94]将跨模态索引任务与表征学习相结合,利用模态内部和模态间的相关关系构造了基于栈式自动编码器的多模态检索模型。但是这种表征方式只是对不同模态的信息进行了浅层的融合,并没有提取出多模态中的有效信息。

相对联合表征,其他表征方法会通过约束和联合训练来实现模态之间的调和,如图 5 所示,通过某种条件和约束使得不同模态的表征达到统一。

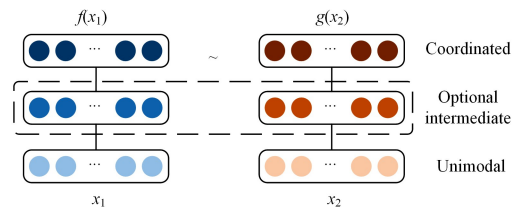


图 5 多模态数据的调和表征

Fig. 5 Coordinated representation of multimodal data

典型的方法如基于图像语义嵌入的表征方法(Visual-Semantic Embedding, VSE)^[95-96]通过对象在不同模态上的相似度来构造排序损失函数,从而使得两个模态(图像和文本)的表征能够产出一致的相似度。根据这个方法,排序嵌入(Order-Embedding)^[88]改进了 VSE 中的全局排序,采用偏序的方法统一两个模态的信息。文献[97]也通过改进损失函数来获得更好的多模态学习性能。除了通过相似度来联合两个模态,文献[98]将点击数据引入多模态的表征学习,通过用户对图像和文本的点击行为将图像和文本构建成一个网络图,利用随机游走学习图像-文本二元组的表征。在文献[99-100]中,监督信息被用来统一多模态,以学习模态之间的一致性表征。现有的多模态数据表征方法更多地关注模态之间的一致性信息,而很少关注模态之间的互补信息。

4 表征学习的未来研究方向

随着复杂异构数据的不断产生和不同应用任务的需求变化,对数据表征的要求也不断提高,同时数据表征也将越来越深刻地影响机器学习的发展速度和实际应用场景。结合复杂异构数据的数据特点和现有机器学习算法的发展方向,现列出以下4个未来的研究方向以供探讨。

4.1 通用表征学习方法与定制化的表征学习算法相结合

最早的数据表征都是采用通用的方法,例如独热表征,但是随着数据复杂性的增大,这种通用的表征给后续处理算法带来了很大的负担,因此我们希望能够能够在表征学习阶段就将数据内部的一些重要特性放大,同时尽量减少无关和干扰的信息。表征学习过程就是不断加入先验知识的过程,随着先验知识的增多或者强化,数据表征的表达力会越来越强,但也会使表征的泛化能力变弱。同时,我们又不希望为每个应用甚至每种算法都设计不同的表征学习模型,这样就丧失了将表征学习独立出算法的意义了。因此如何保证表征学习模型的通用性和泛化能力,又尽量多地保留有用信息从而帮助后续算法的学习,将成为表征学习的一个重要研究问题。但是,如何突破数据之间的异构性,设计针对多种异构数据的表征学习框架仍然需要进一步的研究。

4.2 具有解释性的表征学习方法

最新的表征学习方法都是基于深度学习产生分布式表征,虽然深度学习模型能够从数据中提取更复杂的关系和更抽象的特征,从而生成表达能力更强的数据表征,但是与原始数据或简单数据表征相比缺乏可解释性,不容易被人理解。如果我们能够学习出具有可解释性的表征,那么也会相应提高后续算法的可解释性。

我们可以直接采用具有可解释性的模型,如基于规则的方法(如决策树)、基于单个特征的方法(如线性回归)和基于实例的方法(如聚类)等。但是具有可解释性的方法往往是一些经典的简单方法,若仅仅采用这些经典方法可能无法提取复杂数据中的特征,因此组合多种简单模型或采用集成模型对数据建模不失为一种策略。除了采用本身就具有可解释性的模型,我们还可以在建模之后采用可解释性的方法对表征结果做出解释,这样即可在建模时采用更复杂、更抽象的方法,如深度学习模型。但是,目前的可解释性与模型结构结合得还不够紧密,更多是对结果的解释,还未真正对模型内部与结果的关系进行有效解释,因此还需要进一步的研究。

4.3 融入部分监督信息的表征学习

本文中的大部分表征学习模型都是无监督模型,因为无监督模型的通用性更强,但是设计无监督模型的难度也较大。正如前文所说,表征学习的过程就是引入先验学习的过程。对于无监督学习来说利用的先验知识较少,而且同样的先验知识并不是对所有无监督学习方法都适用,盲目引入先验知识可能会损失数据中的有用信息。特别是对于复杂异构数据,完全采用无监督的表征学习方式可能无法捕获到数据中真正有用的特征。

同时,现实世界中的复杂异构数据很难获取全部标签信息,且数据在不同任务中的标签可能差别较大,采用有监督的

方式进行表征学习的可行性很小,因此,对于复杂异构数据,引入部分监督信息是一种可行的折中办法,这种监督信息不一定是数据的分类标签,可以是来自于不同领域的共识信息。如何让有限的监督信息发挥更大的作用,同时又不影响表征的泛化性,也是未来一个很有意义的研究方向。

4.4 多源异构数据表征中共识信息和互补信息的提取

相比单一数据源的数据,多源异构数据的表征学习不仅要面对数据内部的复杂关系,还需要应对多源数据之间的异构性和融合问题。大部分基于多源数据的学习算法,如多视角学习,都重点关注多个数据视图之间的共识信息,而对互补或者矛盾的信息缺乏关注。而我们希望数据表征能够融合、统一多个数据源的数据,除了克服数据之间的异构性,还能够从多个数据源中吸取互补信息,从而使得融合后的表征能够拥有比单一数据源更丰富、更有效的信息。但是其面临一系列待进一步研究的问题,例如,如何定义互补信息、如何处理数据源之间的矛盾信息等。解决这些问题不仅需要借助任务目标,还可能需要监督信息的指导,因此这个方向与融入监督信息的表征学习也紧密相关。

结束语 本文对多种复杂异构数据的表征学习现状展开了讨论,对多种典型数据的表征学习模型进行了全面的综述;分别对离散数据、网络数据、文本和图像数据的代表性表征学习模型进行了介绍,并详细分析了现有的针对混合数据、属性网络数据、跨领域和多模态数据的表征学习模型和其优缺点;最后对复杂异构数据表征学习的未来发展方向进行了展望。

参 考 文 献

- [1] 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.
- [2] HARRIS Z S. Distributional structure[J]. Word, 1954, 10(2/3): 146-162.
- [3] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[J]. arXiv: 1301.3781, 2013.
- [4] GROVER A, LESKOVEC J. node2vec: Scalable feature learning for networks[C]// Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2016: 855-864.
- [5] BENJIO Y. Deep learning of representations for unsupervised and transfer learning[C]// Proceedings of ICML Workshop on Unsupervised and Transfer Learning. 2012: 17-36.
- [6] GOODFELLOW I J, COURVILLE A, BENGIO Y. Spike-and-slab sparse coding for unsupervised feature discovery[J]. arXiv: 1201.3382, 2012.
- [7] COHEN P, WEST S G, AIKEN L S. Applied multiple regression/correlation analysis for the behavioral sciences[M]. Psychology Press, 2014.
- [8] BENJIO Y, LECUN Y. Scaling learning algorithms towards AI[J]. Large-scale Kernel Machines, 2007, 34(5): 1-41.
- [9] JOLLIFFE I. Principal component analysis[M]. Springer Berlin Heidelberg, 2011.

- [10] AIZAWA A. An information-theoretic perspective of tf-idf measures[J]. *Information Processing & Management*, 2003, 39(1): 45-65.
- [11] AHMAD A, DEY L. A method to compute distance between two categorical values of same attribute in unsupervised learning for categorical data set[J]. *Pattern Recognition Letters*, 2007, 28(1): 110-118.
- [12] IENCO D, PENSA R G, MEO R. From context to distance: Learning dissimilarity for categorical data clustering[J]. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 2012, 6(1): 1.
- [13] JIA H, CHEUNG Y, LIU J. A new distance metric for unsupervised learning of categorical data [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2015, 27(5): 1065-1079.
- [14] WANG C, CHI C H, ZHOU W, et al. Coupled interdependent attribute analysis on mixed data [C] // *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.
- [15] JIAN S, CAO L, PANG G, et al. Embedding-based Representation of Categorical Data by Hierarchical Value Coupling Learning [C] // *IJCAL*. 2017: 1937-1943.
- [16] JIAN S, PANG G, CAO L, et al. CURE: Flexible Categorical Data Representation by Hierarchical Coupling Learning [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2018, 31(5): 853-866.
- [17] ZHANG K, WANG Q, CHEN Z, et al. From categorical to numerical: Multiple transitive distance learning and embedding [C] // *Proceedings of the 2015 SIAM International Conference on Data Mining*. Society for Industrial and Applied Mathematics, 2015: 46-54.
- [18] BALASUBRAMANIAN M, SCHWARTZ E L. The isomap algorithm and topological stability[J]. *Science*, 2002, 295(5552): 7.
- [19] YIN M, GAO J, LIN Z. Laplacian regularized low-rank representation and its applications [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 38(3): 504-517.
- [20] HE X, NIYOGI P. Locality preserving projections [C] // *Advances in Neural Information Processing Systems*. 2004: 153-160.
- [21] CAO S, LU W, XU Q. Grarep: Learning graph representations with global structural information [C] // *Proceedings of the 24th ACM International Conference on Information and Knowledge Management*. ACM, 2015: 891-900.
- [22] PEROZZI B, AIRFOU R, SKIENA S. Deepwalk: Online learning of social representations [C] // *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2014: 701-710.
- [23] TANG J, QU M, WANG M, et al. Line: Large-scale information network embedding [C] // *Proceedings of the 24th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 2015: 1067-1077.
- [24] WANG D, CUI P, ZHU W. Structural deep network embedding [C] // *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2016: 1225-1234.
- [25] CAO S, LU W, XU Q. Deep neural networks for learning graph representations [C] // *Thirtieth AAAI Conference on Artificial Intelligence*. 2016.
- [26] CUI P, WANG X, PEI J, et al. A survey on network embedding [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2018, 31(5): 833-852.
- [27] CAI H, ZHENG V W, CHANG K C C. A comprehensive survey of graph embedding: Problems, techniques, and applications [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2018, 30(9): 1616-1637.
- [28] DEERWESTER S, DUMAIS S T, FURANAS G W, et al. Indexing by latent semantic analysis [J]. *Journal of the American Society for Information Science*, 1990, 41(6): 391-407.
- [29] BLEI D M, NG A Y, JORDAN M I. Latent dirichlet allocation [J]. *Journal of Machine Learning Research*, 2003, 3(1): 993-1022.
- [30] HOFMANN T. Probabilistic latent semantic indexing [C] // *ACM SIGIR Forum*. ACM, 2017, 51(2): 211-218.
- [31] WILSON A T, CHEW P A. Term weighting schemes for latent dirichlet allocation [C] // *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, 2010: 465-473.
- [32] BENGIO Y, DUCHARME R, VINCENT P, et al. A neural probabilistic language model [J]. *Journal of Machine Learning Research*, 2003, 3(Feb): 1137-1155.
- [33] PENNINGTON J, SOCHER R, MANNING C. Glove: Global vectors for word representation [C] // *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2014: 1532-1543.
- [34] PETERS M E, NEUMANN M, IYYER M, et al. Deep contextualized word representations [J]. *arXiv:1802.05365*, 2018.
- [35] DEVLIN J, CHANG M W, LEE K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding [J]. *arXiv:1810.04805*, 2018.
- [36] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C] // *Advances in Neural Information Processing Systems*. 2017: 5998-6008.
- [37] YANG Z, DAI Z, YANG Y, et al. XLNet: Generalized Autoregressive Pretraining for Language Understanding [J]. *arXiv:1906.08237*.
- [38] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition [J]. *Proceedings of the IEEE*, 1998, 86(11): 2278-2324.
- [39] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks [C] // *Advances in Neural Information Processing Systems*. 2012: 1097-1105.
- [40] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. *arXiv:1409.1556*, 2014.
- [41] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015: 1-9.

- [42] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016;2818-2826.
- [43] SZEGEDY C, IOFFE S, VANHOUCKE V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning[C]// Thirty-First AAAI Conference on Artificial Intelligence. 2017.
- [44] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016;770-778.
- [45] AL-SAFFAR A A M, TAO H, TALAB M A. Review of deep convolution neural network in image classification[C]// 2017 International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET). IEEE, 2017;26-31.
- [46] CHEN C, QI F. Review on Development of Convolutional Neural Network and Its Application in Computer Vision[J]. Journal of Computer Science, 2019, 46(3):63-73.
- [47] DAVID G, AVERBUCH A. SpectralCAT: Categorical spectral clustering of numerical and nominal data[J]. Pattern Recognition, 2012, 45(1):416-433.
- [48] CAO L. Coupling learning of complex interactions[J]. Information Processing & Management, 2015, 51(2):167-186.
- [49] CAO L, OU Y, PHILIP S Y. Coupled behavior analysis with applications[J]. IEEE Transactions on Knowledge and Data Engineering, 2011, 24(8):1378-1392.
- [50] WEI M, CHOW T, CHAN R. Clustering heterogeneous data with k-means by mutual information-based unsupervised feature transformation[J]. Entropy, 2015, 17(3):1535-1548.
- [51] HUANG Z. Clustering large data sets with mixed numeric and categorical values[C]// Proceedings of the 1st Pacific-asia Conference on Knowledge Discovery and Data Mining (PAKDD). 1997;21-34.
- [52] CHEN J Y, HE H H. A fast density-based data stream clustering algorithm with cluster centers self-determined for mixed data[J]. Information Sciences, 2016, 345:271-293.
- [53] JI J, PANG W, ZHOU C, et al. A fuzzy k-prototype clustering algorithm for mixed numeric and categorical data[J]. Knowledge-Based Systems, 2012, 30:129-135.
- [54] JIA H, CHEUNG Y M. Subspace clustering of categorical and numerical data with an unknown number of clusters[J]. IEEE Transactions on Neural Networks and Learning Systems, 2017, 29(8):3308-3325.
- [55] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks [J]. Science, 2006, 313(5786):504-507.
- [56] BALDI P. Autoencoders, unsupervised learning, and deep architectures[C]// Proceedings of ICML Workshop on Unsupervised and Transfer Learning. 2012;37-49.
- [57] YANG L, JIN R. Distance metric learning: A comprehensive survey[J]. Michigan State University, 2006, 2(2):4.
- [58] FROME A, SINGER Y, SHA F, et al. Learning globally-consistent local distance functions for shape-based image retrieval and classification[C]// 2007 IEEE 11th International Conference on Computer Vision. IEEE, 2007:1-8.
- [59] CHECHIK G, SHARMA V, SHALIT U, et al. Large scale online learning of image similarity through ranking[J]. Journal of Machine Learning Research, 2010, 11(3):1109-1135.
- [60] JIAN S, HU L, CAO L, et al. Metric-based auto-instructor for learning mixed data representation[C]// Thirty-Second AAAI Conference on Artificial Intelligence. 2018.
- [61] LI J, ZHU J, ZHANG B. Discriminative deep random walk for network classification[C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. 2016, 1:1004-1013.
- [62] TU C, ZHANG W, LIU Z, et al. Max-margin deepwalk: Discriminative learning of network representation[C]// IJCAI. 2016;3889-3895.
- [63] PAN S, WU J, ZHU X, et al. Tri-party deep network representation[C]// Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence. AAAI Press, 2016:1895-1901.
- [64] DAI H, DAI B, SONG L. Discriminative embeddings of latent variable models for structured data[C]// International Conference on Machine Learning. 2016;2702-2711.
- [65] WEI X, XU L, CAO B, et al. Cross view link prediction by learning noise-resilient representation consensus[C]// Proceedings of the 26th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2017;1611-1619.
- [66] NIEPERT M, AHMED M, KUTZKOV K. Learning convolutional neural networks for graphs[C]// International Conference on Machine Learning. 2016;2014-2023.
- [67] ZHANG D, YIN J, ZHU X, et al. User Profile Preserving Social Network Embedding[C]// IJCAI. 2017;3378-3384.
- [68] YANG C, LIU Z, ZHAO D, et al. Network representation learning with rich text information[C]// Twenty-Fourth International Joint Conference on Artificial Intelligence. 2015.
- [69] ZHANG Z, YANG H, BU J, et al. ANRL: Attributed Network Representation Learning via Deep Neural Networks[C]// IJCAI. 2018;3155-3161.
- [70] TU C, LIU H, LIU Z, et al. Cane: Context-aware network embedding for relation modeling[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. 2017;1722-1731.
- [71] GAO H, HUANG H. Deep Attributed Network Embedding [C]// IJCAI. 2018;3364-3370.
- [72] JIAN S, HU L, CAO L, et al. Evolutionarily learning multi-aspect interactions and influences from network structure and node content[C]// AAAI-19. 2019.
- [73] CHANG S, HAN W, TANG J, et al. Heterogeneous network embedding via deep architectures[C]// Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2015;119-128.
- [74] ALHARBI B, ZHANG X. Learning from your network of friends: a trajectory representation learning model based on online social ties[C]// 2016 IEEE 16th International Conference on

Data Mining (ICDM). IEEE,2016:781-786.

- [75] ZHANG Q,WANG H. Not all links are created equal: An adaptive embedding approach for social personalized ranking[C]// Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2016:917-920.
- [76] ODENA A,OLAH C,SHLENS J. Conditional image synthesis with auxiliary classifier gans[C]// Proceedings of the 34th International Conference on Machine Learning. JMLR,2017:2642-2651.
- [77] HIGGINS I,MATTHEY L,PAL A,et al. Beta-vae: Learning basic visual concepts with a constrained variational framework [C]// International Conference on Learning Representations. 2017,3.
- [78] CHEN X,DUAN Y,HOUTHOOFT R,et al. Infogan: Interpretable representation learning by information maximizing generative adversarial nets[C]// Advances in Neural Information Processing Systems. 2016:2172-2180.
- [79] GOODFELLOW I,POUGET-ABADIE J,MIRZA M,et al. Generative adversarial nets[C]// Advances in Neural Information Processing Systems. 2014:2672-2680.
- [80] RADFORD A,METZ L,CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. arXiv:1511.06434,2015.
- [81] REZENDE D J,MOHAMED S,WIERSTRA D. Stochastic Backpropagation and Approximate Inference in Deep Generative Models [C] // International Conference on Machine Learning. 2014:1278-1286.
- [82] LIU Y C,YEH Y Y,FU T C,et al. Detach and adapt: Learning cross-domain disentangled deep representation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018:8867-8876.
- [83] GONZALEZ-GARCIA A,VAN DE WEIJER J , BENGIO Y. Image-to-image translation for cross-domain disentanglement[C]// Advances in Neural Information Processing Systems. 2018: 1287-1298.
- [84] LIU A H,LIU Y C,YEH Y Y,et al. A unified feature disentangler for multi-domain image translation and manipulation[C]// Advances in Neural Information Processing Systems. 2018: 2590-2599.
- [85] JIANS,HU L,CAO L,et al. Representation Learning with Multiple Lipschitz-constrained Alignments on Partially-labeled Cross-domain Data [C]// 34th AAAI Conference on Artificial Intelligence. 2020.
- [86] BALTRUŠAITIS T,AHUJA C,MORENCY L P. Multimodal machine learning: A survey and taxonomy[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 41(2):423-443.
- [87] TAIGMAN Y,POLYAK A,WOLF L. Unsupervised cross-domain image generation[J]. arXiv:1611.02200,2016.
- [88] VENDROV I,KIROS R,FIDLER S,et al. Order-embeddings of images and language[J]. arXiv:1511.06361,2015.
- [89] ANTOL S,AGRAWAL A,LU J,et al. Vqa: Visual question answering[C]// Proceedings of the IEEE International Conference on Computer Vision. 2015:2425-2433.
- [90] OUYANG W,CHU X,WANG X. Multi-source deep learning for human pose estimation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 2329-2336.
- [91] ZHANG H,HU Z,DENG Y,et al. Learning concept taxonomies from multi-modal data[J]. arXiv:1606.09239,2016.
- [92] ZHANG J,PENG Y,YUAN M. Unsupervised generative adversarial cross-modal hashing [C] // Thirty-Second AAAI Conference on Artificial Intelligence. 2018.
- [93] VUKOTIĆ V,RAYMOND C,GRAVIER G. Bidirectional joint representation learning with symmetrical deep neural networks for multimodal and crossmodal applications[C]// Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval. ACM,2016:343-346.
- [94] WANG W,OOI B C,YANG X,et al. Effective multi-modal retrieval based on stacked auto-encoders[J]. Proceedings of the VLDB Endowment,2014,7(8):649-660.
- [95] FROME A,CORRADO G S,SHLENS J,et al. Devise: A deep visual-semantic embedding model[C]// Advances in Neural Information Processing Systems. 2013:2121-2129.
- [96] KIROS R,SALAKHUTDINOV R,ZEMEL R S. Unifying visual-semantic embeddings with multimodal neural language models[J]. arXiv:1411.2539,2014.
- [97] WANG L,LI Y,HUANG J,et al. Learning two-branch neural networks for image-text matching tasks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2018,41(2):394-407.
- [98] WU F,LU X,SONG J,et al. Learning of multimodal representations with random walks on the click graph[J]. IEEE Transactions on Image Processing,2015,25(2):630-642.
- [99] KANG C,XIANG S,LIAO S,et al. Learning consistent feature representation for cross-modal multimedia retrieval [J]. IEEE Transactions on Multimedia,2015,17(3):370-381.
- [100] NIAN F,BAO B K,LI T,et al. Multi-Modal Knowledge Representation Learning via Webly-Supervised Relationships Mining [C]// Proceedings of the 25th ACM international conference on Multimedia. ACM,2017:411-419.



JIAN Song-lei, born in 1991, Ph.D, assistant research fellow, is member of China Computer Federation (CCF). Her main research interests include representation learning, machine learning and complex network analysis.



LU Kai, born in 1973, research fellow, Ph.D supervisor, is member of China Computer Federation (CCF). His main research interests include parallel and distributed system software, operating systems and machine learning.