

The Research of Image Semantic Analysis Technology and Application Based on Knowledge Graph

Liqiong Deng, Guixin Zhang, Xiangning Hao

Department of Command Information System and Network, Air Force Communication NCO Academy, Dalian Liaoning

Email: tigerss1016@163.com, singlet@163.com, hao_xiangning@163.com

Received: Aug. 24th, 2018; accepted: Sep. 6th, 2018; published: Sep. 13th, 2018

Abstract

The semantic analysis technology of image has always been a difficult point in image field. As an intelligent and efficient way of organizing, knowledge graph can help users accurately query the information. This paper firstly puts forward an image semantic analysis process based on knowledge graph, then adopts deep learning model to describe image's features. Image semantic knowledge fusion and processing is studied on this basis; a multilevel image semantic model constructed has ability to manage entity triples and support automatic construction. Finally, applications in semantic retrieval, association and visualization are analyzed, which has some guiding significance for information organization and knowledge management of media semantic.

Keywords

Knowledge Graph, Deep Learning, Image, Semantic Analysis, Semantic Retrieval

基于知识图谱的图像语义分析技术及应用研究

邓莉琼, 张贵新, 郝向宁

空军通信士官学校指挥信息系统与网络系, 辽宁 大连

Email: tigerss1016@163.com, singlet@163.com, hao_xiangning@163.com

收稿日期: 2018年8月24日; 录用日期: 2018年9月6日; 发布日期: 2018年9月13日

摘要

图像的语义分析技术一直是图像领域的研究难点之一, 知识图谱作为一种智能的知识组织方式, 可以帮

助用户迅速、准确地查询到所需要的信息。本文首先提出了一种基于知识图谱的图像语义分析流程，然后采用了深度表达模型对图像的机构化语义信息进行描述和抽取，在此基础上研究了基于知识图谱的图像语义知识融合和加工技术，构建后的多层次图像语义模型具备管理实体关系三元组的能力、支持图谱的自动构建与多模式查询。最后基于该思路分析了图像语义分析技术在语义检索、关联分析及可视化方面等的应用，对媒体语义中的信息组织和知识管理有一定的指导意义。

关键词

知识图谱，深度学习，图像，语义分析，语义检索

Copyright © 2018 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

图像作为一种内容丰富、表现直观的媒体信息，越来越受到人们的青睐，如何有效地描述和检索这些图像信息成为研究者们所关心的热点问题。尤其是随着近几年深度学习技术的日趋完善，传统的计算机视觉任务得到了前所未有的发展，提升了图像信息提取的能力。然而图像语义理解和语义分析等研究领域还停留在传统的图像信息抽取和图像标注技术上。

由于知识是对信息的进一步组织和抽象，符合人类活动的语义和逻辑。知识图谱技术在增进信息的组织、管理和理解领域具有巨大的应用潜力，是对图像视觉语义理解的一个行之有效的途径。相对于信息，知识能更直接地指导人的决策和行动。因此本文提出了一种基于知识图谱的图像语义分析的思路，将知识图谱技术用于构建图像的语义概念关系之中，能将完整的图像内容转换成可直观理解的结构化语言表达及可视化分析，在图像理解中起着至关重要的作用，同时还为基于语义的图像检索和多媒体关联提供了一定的解决思路。

2. 相关工作

图像语义分析的难点主要集中在两方面：1) 图像的视觉表达和语义之间很难建立合理关联，描述实体间产生巨大的语义鸿沟(Semantic Gap)；2) 语义本身具有表达的多义性和不确定性。目前，越来越多的研究在关注上述瓶颈，并致力于有效模型和方法以实现图像理解中的语义表达。目前的解决思路可大致分为三类，第一条思路侧重于图像本身的研究，通过构建和图像内容相一致的模型或方法，将语义隐式地融入其中，建立文本-图像的有向联系，核心在如何将语义融于模型和方法中；第二条思路从语义本身的句法表达和结构关系入手，分析其组成及相互关系，通过建立与之类似的图像视觉元素结构表达，将语义描述和分析方法显式地植入包含句法关系的视觉图中，建立图像-文本的有向联系；第三条思路面向应用，以基于内容图像检索为核心，增加语义词汇规模，构建多语义多用户多进程的图像检索查询系统。

图像的语义分析是建立在图像的信息抽取技术上的，图像信息抽取研究经历了三个阶段：1) 利用文本来描述图像特征；2) 图像底层视觉特征；3) 图像的语义内在特征。早期的图像信息抽取是利用底层图像特征，如方向梯度直方图 HOG 和尺度不变特征转换 SIFT [1]。基于机器学习的方法从最简单的像素级阈值法、基于像素聚类的分割方法到基于图论划分的分割方法[2]。基于机器学习的方法，先将输入图像分为一些独立的区域块，并提取每个区域块的特征，然后根据一定的规则建立图像特征与语义类别之

间的概率模型，建立起能量函数，并通过手工标注的特征库，迭代计算对能量函数进行优化，得到最优的参数，最终得到图像信息模型。这种机器学习的方法过于依赖手工标注的特征库，难以广泛表示图像特征，在实际应用中有很大的局限性。

面对海量的图像信息，人们期望以更加智能的方式组织图像资源。为了满足这种需求，知识图谱应运而生。它们力求通过将知识进行更加有序、有机的组织，对用户提供更加智能的访问接口，使用户可以更加快速、准确地访问自己需要的知识信息，并进行一定的知识挖掘和智能决策[3]。知识图谱技术的出现使得信息可以在语义层面上进行整合，这种语义层次的关联技术能够为图像的语义分析研判提供强有力的支撑。知识图谱即为用图对知识和知识间关系进行建模。知识图谱的功能主要体现在知识组织、展示与搜索方面[4]：第一，在一定程度上克服自然语言的歧义性；第二，把经过梳理、总结的知识提供给用户；第三，提供更深入更广阔的知识，知识图谱尝试通过对其他用户相关的搜索记录进行推理，激发用户对知识的搜索兴趣，从而进行一次全新的查询操作。

3. 基于知识图谱的图像语义分析流程

针对图像的语义分析，本文所提出的基于知识图像的图像语义分析流程图如图 1 所示，分析流程包括数据获取、信息抽取、知识融合加工、知识存储、知识应用和可视化结果 6 个部分。

图像数据获取是采集原始的图像数据，即可以包含结构化数据，如标注文本等信息，也可以包含半结构化和非结构化的图像特征信息等；图像的信息抽取是将图像数据进行抽象归纳为更具有语义特点的单元，例如实体抽取、实体之间的关系抽取以及实体的属性抽取等，为后续的知识提取做铺垫；图像的知识融合和加工是在图像信息抽取的基础上将信息升级为知识，例如具有相同表达但不同信息的实体消歧、具有相同意思但不同表达的知识合并、对知识进行概念归纳的本体构建以及丰富语义内涵的知识推理；图像的知识存储是利用知识图谱的三元组表达方式，结合前面两步获取的知识，将其存储在数据库中并进行知识更新；可视化结果则是利用知识图谱的可视化工具将图像的知识图谱语义分析记过以网状可视化方式直观的展现出图像知识组成。

4. 基于深度表达模型的图像结构化语义信息抽取技术

图像信息抽取是通过一定的方法将图像中的每个像素分为不同的语义类别，最终得到不同的实体、关系、属性等信息。传统的图像语义特征的提取是以图像底层视觉特征为基础的，即首先通过相关的图像特征提取方法提取出图像的颜色、纹理、形状、轮廓等底层特征信息，然后寻找图像底层特征与高层语义的相关性，最后建立底层特征与高层语义的映射关系[5]。

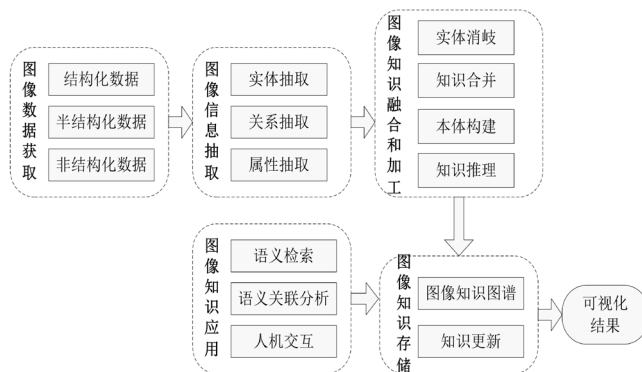


Figure 1. Image semantic analysis flow based on knowledge graph
图 1. 基于知识图谱的图像语义分析流程

近年来深度卷积神经网络(DCNNs) [6]的发展,使得计算机视觉任务中的许多问题得到了巨大的突破。由于深度卷积网络可以从大量样本数据中提取出图像的特征,比手工标注特征更好,在图像分类和物体检测等高层计算机视觉任务上获得了巨大的成功。LinG 等[7]通过多尺度图像作为输入产生不同尺度的特征图,应用 DCNN 的图像信息抽取模型,可以通过大量的样本自主学习特征,打破了手工标注特征的局限性。本文首先提出了面向知识图谱构建的图像语义层次描述框架,然后利用深度学习技术实现图像的主要信息提取。

4.1. 图像语义层次描述框架

该框架主要包括四个层次,第一层关注整体感官信息,着重描述图像中整体的基础特征。例如,图像的纹理、色彩、形状、空间分布;第二层从整体转移至局部,开始关注图像中的实体与场景。实体的判断主要包括人物、动物、物体等。场景的信息抽取是交代图像整体的语义背景,同时提供依赖情境,是增加图像纵向关联不可缺少的基础数据。

第三层聚焦于实体的关系和属性上,主要是识别其与其他实体之间的交互位置关系。交互位置是立足于实体相对位置的描述,判断实体关系及实体交互位置关系是图像语义关联中的重要组成部分。实体的属性是描述实体包含的具体参数,例如种类、大小、颜色等信息,识别在该情境下各个实体的内涵。最后建立在属性与交互位置关系之上的行为交互语义。充分挖掘实体之间的横向关联和纵向关联;第四层是一种整体语义的回归,是判断完所有局部和聚焦信息后的语义概括,是全部对象行为的语义整合(图2)。

4.2. 实体抽取

人类视觉系统对复杂的场景结构具有较强的感知能力,例如场景中的物体类别、物体的空间位置、物体间的组合关系等。如何将上述具有明显结构的语义信息融入到深度模型的设计中,是视觉研究领域广泛关注的问题。神经网络的出现,使端到端的图像处理成为可能,而卷积神经网络(CNN)便是深度学习与图像处理技术相结合所产生的经典模型。本文采取的具体方法是文献[8]所提出的融合结构化语义的深度表达模型,即 CNN-RNN 的混合神经网络图像语义表达。该模型把一系列带有结构的语法知识融入到深度表达学习中,借助递归神经网络的特性对图像中的物体、物体间的关系等进行结构化的组合与表达。最终模型实现图像的深度层次化语义解析。

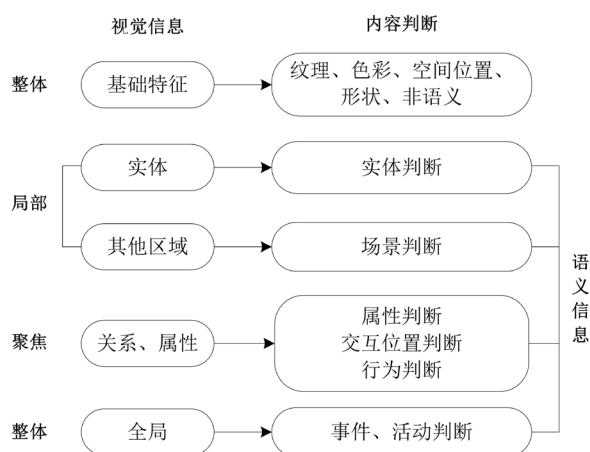


Figure 2. Image semantic hierarchy framework
图2. 图像语义层次框架

该网络包含了 2 个相互连接的神经网络结构，底层是深度卷积神经网络(CNN) [9]，上层是递归神经网络(RNN)。在 CNN-RNN 混合模型中、CNN 模型和 RNN 模型将协同工作。其中，CNN 模型能够生成具有判别能力的图像特征表达，被广泛地应用于图像分类与物体识别任务中，因此将其用作物体类别的表达学习；RNN 模型则被一系列工作[10]证明能够对图像或者自然语言中的结构化组合关系进行预测，其原因在于该类网络能够同时对语义和结构化表达进行递归地学习，故而利用 RNN 模型把 CNN 模型产生的特征表达作为输入，用于进一步生成场景的结构化配置。该算法经实验验证，对图像实体语义抽取的准确度较高，具体算法实现过程如图 3 所示，由于该算法不是本文关注主要问题，不在文中详细阐述。

4.3. 关系和属性抽取

在 CNN-RNN 混合模型中，RNN 模型将利用 CNN 模型所输出的每个语义类别的特征表达来生成图像的解析树，从而对图像进行内容上的分层表达，同时预测语义实体之间的交互关系。这里的 RNN 模型包含 5 个子网络，分别是中间转换网络、节点合并网络、语义解释网络、关系分类网络和合并分数网络。

在关系和属性抽取时，第一种方法是基于词法模式的原理根据语句构成成分之间的语义关系来预测语句整体的意义。比如通过对文本句法的分析，对于一个以动词为核心的短句，可以抽取出实体之间的潜在关系；另外一种是其现分析方法，这是一种定量与定性相结合的分析方法，其具体步骤为，先将待处理文本转化为数字形式表达的信息，然后使用不同的数学方法对文本进行定量计算和分析，最后结合定性分析的结果对文本中的分类关系进行综合分析；最后，还有基于开放链接数据和基于在线百科的方法，该方法通过百科类等网站规则的知识分类体系，定义或者学习知识分类的规则和特征，从而对隐藏的分类关系和属性进行准确地抽取。

5. 图像语义知识融合和加工技术

通过信息抽取获取的这些结果中可能包含大量的冗余和错误信息，有必要对其进行清理和整合。

5.1. 知识合并

为了剔除冗余信息，首先需要进行实体对齐与消歧。实体对齐是知识图谱构建过程中的重要步骤之一，通过实体对齐，将同一个知识图谱内部的实体进行精简，从而实现知识图谱之间的链接与合并，由此构建一个规模更大，服务范围更广泛的知识图谱系统。实体对齐是对于物理世界中的同一个对象，要识别出它在不同语言，不同地域，不同数据源下不同的表示形式，之后用一个全局唯一的编号来标识。实体对齐算法的主要过程是根据具体的知识图谱的特点和处理方法，利用不同的实体识别方法，例如传统概率模型的方法、以及机器学习的方法，来完成实体对齐的任务[11]。

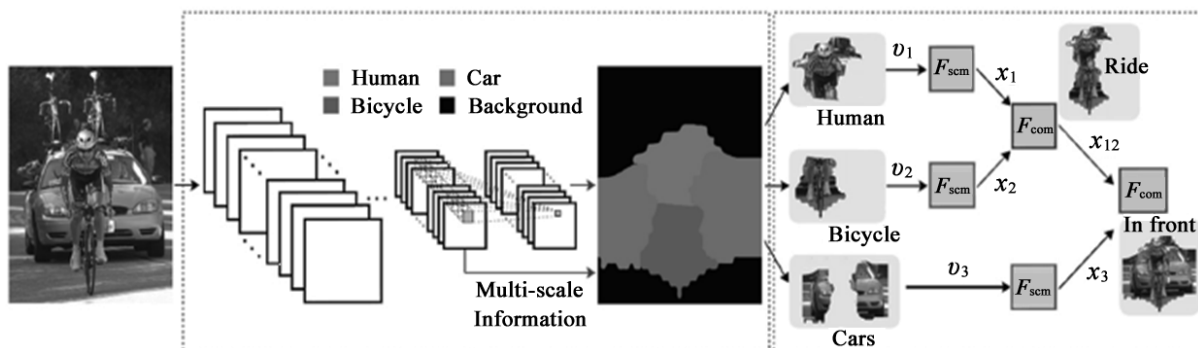


Figure 3. Image structured semantic realized process based-on CNN-RNN model
图 3. 基于 CNN-RNN 模型的图像结构化语义实现流程

实体消歧是专门用于解决同名实体产生歧义问题的技术。通过实体消歧, 就可以根据当前的语境, 准确建立实体链接。同义关系是指在概念层面上相同或相似的实体。同义关系抽取的目标是寻找那些字面不同但是指代同一概念、实体或属性的术语。传统的基于模板的同义关系抽取方法灵活性不够, 模板的覆盖率不高, 导致该方法的正确率和召回率都较低。孙霞等人[12]提出了一种自动抽取同义关系的机器学习算法。该方法具有领域自适应性, 可以将训练好的分类器应用到不同的领域文本中, 与基于模板的方法相比较, 抽取结果的精度有了普遍提高。随着统计语言模型和深度算法的逐渐成熟, 对于寻找自然语言中的同义关系的任务通常可以取得较满意的正确率。

5.2. 本体构建

在图像语义领域中, 本体在某种意义上就是知识模型。图像语义的本体构建就是在领域专家的帮助下, 建立基于领域概念知识的领域本体, 收集信息源中的图像, 并参照已建立的领域本体, 把收集来的图像按规定的格式存储在元数据库(关系数据库、知识库等)中。图像本体是对图像的一种描述模式, 任何一幅图像都是图像本体的一个实例, 并可以用本体语言形式化地表示。

对于图像领域本体的构建, 本文采用形式概念分析和 Word2vec 神经网络工具, 结合结构化和非结构化资源, 利用结构化资源概念层次明确、非结构化资源语义丰富以及容易获取的特点, 有效降低领域图像本体构建的复杂度, 节省时间人力成本, 能提高图像领域本体的构建效率, 挖掘出隐含的概念以及概念间关系, 丰富图像领域本体的语义信息, 能较好地反映图像领域的知识结构[13]。

5.3. 知识演化推理

图像中语义的演化与上文提出的基于表示过程的层次描述框架是密不可分的(见图 2)。基于表示过程的层次框架本身是面向于用户解读过程, 是一个语义抽象层次不断上升的过程。在该过程中, 由于被添加新的解读信息, 下层的语义将演化成更具意义的新内容。其次, 语义的演化需要各种视觉信息, 视觉信息在理解加工后才会具有更符合人类认知的概念和语义表达。

知识推理包含对实体属性的推理和对实体关系的推理。实体属性的推理主要是对会发生变化的实体属性值进行实时发现、推理、更新或者为实体创建新的属性; 实体之间关系的推理是对实体之间潜在的关系进行推断和扩充。知识的推理方法可以分为基于逻辑的推理和基于图的推理。基于逻辑的推理主要包括一阶谓词逻辑、描述逻辑以及基于规则的推理, 推理规则包含针对实体属性的规则和针对实体关系的规则; 基于图的推理方法主要有基于神经网络模型或路径规划算法。

6. 图像知识图谱的应用

6.1. 语义检索

基于语义的检索对于克服图像信息中的语义鸿沟具有重要的作用, 基于知识图谱生成的图像语义框架可以更好的服务于语义检索领域, 这是由于与传统的基于关键字匹配的搜索引擎工作原理不同的是, 知识图谱利用概念、实体的匹配度返回给用户与搜索相关的更全面的知识体系。

语义检索是基于之前的语义组织体系, 实现知识关联和概念语义检索的智能化检索方式。知识图谱中的语义检索包含两类核心任务: 一是利用相关性在知识库中找到相应的实体; 二是在此基础上根据实体的类别、关系及相关性等信息找到关联的实体[14]。通过对知识库进行深层次的知识挖掘与提炼后, 检索系统为用户反馈出具有重要性排序的准确且完整的知识, 并推荐用户感兴趣的相关知识。

6.2. 语义关联分析

语义关联分析的基本任务是根据主题、形式、自然属性、社会属性等, 链接具有相似语义信息的图

像等视觉媒体。在各种跨媒体关联类型中最关键的是关联数据模型。传统的数据模型着重解决的是数据个体的存储组织与管理问题,而关联数据模型不仅要能够存储组织和管理非结构化媒体数据,还要能够存储组织和管理媒体数据之间的逻辑关联,能够存储组织和管理每个媒体数据的多维属性以及每个关联的多维属性,能够反映跨媒体数据及其关联的时空属性、语义涵义、权重、概率等特性。知识图谱为基础构建数据模型,能够更好地实现传统数据模型所不能支持的多种智能分析,时空关联分析、逻辑关联分析、语义相似性搜索、数据世系管理与分析、数据溯源与核查等,提升多媒体信息之间关联分析能力。

6.3. 知识可视化

知识图谱可视化将知识库中的信息转化为更方便用户理解的方式进行呈现,用户可以一目了然地了解到他需要的知识;同时提供了更加丰富的文本信息,增加了更多的用户交互元素,提升用户体验,引导用户在短时间内获取到更多的知识。美国德雷克塞尔大学研发出的可视化软件 Citespace 为知识图谱研究提供了极大的便利, Citespace 是基于 JAVA 开发的可视化软件,可以免费下载,使用起来也很便利。Citespace 软件有强大的词汇分析能力,可用于做知识图谱的可视化分析,推荐选用 Citespace5.1.R6 版本。

7. 结论

基于知识图谱技术实现的图像语义分析方法可以使人们更便捷、准确地获取到自己所需要的图像语义信息并增强对图像的理解,具有重大的研究意义和使用价值。在未来信息爆炸的世界中,知识图谱将作为人们访问知识信息的接口,在知识组织和展现上扮演越来越重要的角色。

目前利用知识图谱实现对图像等视觉媒体的语义分析研究还处于初级阶段,仍然存在很多的挑战和难题需要解决,例如图像语义知识库的自动扩展,推理规则的学习等。另外,利用知识图谱跨媒体分析同样是一大挑战。相信本文提出的基于知识图谱的图像语义分析方法将为多媒体语义分析领域提供一定的解决思路和参考价值。

参考文献

- [1] Zhou, H., Yuan, Y. and Shi, C. (2009) Object Tracking Using SIFT Features and Mean Shift. *Computer Vision and Image Understanding*, **113**, 345-352. <https://doi.org/10.1016/j.cviu.2008.08.006>
- [2] Krahenbühl, P. and Koltun, V. (2011) Efficient Inference in Fully Connected Crfs with Gaussian Edge Potentials. *Advances in Neural Information Processing Systems*. MIT Press, Massachusetts, 109-117.
- [3] 曹倩, 赵一鸣. 知识图谱的技术实现流程及相关应用[J]. *情报理论与实践*, 2015, 38(12): 13-18.
- [4] Zhang, X.L., Du, C.L., Li, P.S., et al. (2011) Knowledge Graph Completion via Local Semantic Contexts. *Database Systems for Advanced Applications*, **9642**, 432-446. https://doi.org/10.1007/978-3-319-32025-0_27
- [5] Shotton, J., Winn, J. and Rother, C. (2009) Texton Boost for Image Understanding: Multi-Class Object Recognition and Segmentation by Jointly Modeling Texture, Layout, and Context. *International Journal of Computer Vision*, **81**, 2-3. <https://doi.org/10.1007/s11263-007-0109-1>
- [6] He, K., Zhang, X., Ren, S., et al. (2016) Deep Residual Learning for Image Recognition. *IEEE Proceedings of the Conference on Computer Vision and Pattern Recognition*, Piscataway, 770-778.
- [7] Lin, G., Shen, C., van den Hengel, A., et al. (2016) Efficient Piecewise Training of Deep Structured Models for Semantic Segmentation. *IEEE Proceedings of the Conference on Computer Vision and Pattern Recognition*, Piscataway, 3194-3203.
- [8] 张瑞茂, 彭杰锋, 吴恙, 林惊. 融合语义知识的深度表达学习及在视觉理解中的应用[J]. *计算机研究与发展*, 2017, 54(6): 1251-1266.
- [9] Long, J., Shelhamer, E. and Darrell, T. (2015) Fully Convolutional Networks for Semantic Segmentation. *Proceeding of 2015 IEEE Conference on Computer Vision and Pattern Recognition*, Piscataway, 3431-3440. <https://doi.org/10.1109/CVPR.2015.7298965>
- [10] Sharma, A., Tuzel, O. and Jacobs, D.W. (2015) Deep Hierarchical Parsing for Semantic Segmentation. *Proceeding of*

2015 *IEEE Conference on Computer Vision and Pattern Recognition*, Piscataway, 3128-3137.

<https://doi.org/10.1109/CVPR.2015.7298651>

- [11] 庄严, 李国良, 冯建华. 知识库实体对齐技术综述[J]. 计算机研究与发展, 2016, 53(1): 165-192.
- [12] 孙霞, 董乐红. 基于监督学习的同义关系自动抽取方法[J]. 西北大学学报, 2008, 38(1): 35-39.
- [13] 李跃鹏, 金翠, 及俊川. 基于 Word2vec 的关键词提取算法[J]. 科研信息化技术与应用, 2015(4): 54-59.
- [14] 杨思洛, 韩瑞珍. 知识图谱研究现状及趋势的可视化分析[J]. 情报资料工作, 2012, 33(4): 22-28.

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2161-8801, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: csa@hanspub.org