

文章编号: 2095-2163(2021)02-0029-06

中图分类号: TP391

文献标志码: A

基于视觉的道路场景建模

种玉祥, 梁耀中

(上海工程技术大学 机械与汽车工程学院, 上海 201620)

摘要: 步入 21 世纪, 随着人工智能的发展, 智能车的研究成为一大热点, 而智能汽车研究的基础就是定位问题。目前有 2 种定位方法, 一种是实时定位与建图 (SLAM), 另一种则是基于道路场景表征建模的定位方法, 两者各有所长。本文针对基于视觉的道路场景表征建模定位方法进行了优化与改进。首先, 本文提出了一种对点云处理的方法, 对当前 Z 坐标一定距离内的点云取不同权值, 进行加权投影, 以此来构建道路的二维场景。采用 ORB 特征提取算子提取二维特征, 并采用视觉里程计算法获取车辆运动轨迹信息。构建了轻量级神经网络, 用来检测道路标志特征, 例如车道线、斑马线、道路标志牌等。对二维场景精度差的问题进行补充。

关键词: 视觉 SLAM; GPS; 融合定位; 智能车定位; 轻量级神经网络

Vision-based road scene modeling

CHONG Yuxiang, LIANG Yaozhong

(School of Mechanical and Automotive Engineering, Shanghai University of Engineering Science, Shanghai 201620, China)

[Abstract] In the 21st century, with the development of artificial intelligence, the research of smart cars has become a hot topic, and the basis of smart car research is the positioning problem. There are currently two positioning methods, one is simultaneous localization and mapping (SLAM), the other is a positioning method based on road scene representation modeling. Each method has corresponding advantage. This paper optimizes and improves the vision-based road scene representation modeling and positioning method. First of all, this paper proposes a method for point cloud processing, which takes different weights for the point cloud within a certain distance of the current Z coordinate and performs weighted projection to construct a two-dimensional road scene. The ORB feature extraction operator is used to extract two-dimensional features, and the visual mileage calculation method is used to obtain vehicle trajectory information. A lightweight neural network is constructed to detect road sign features, such as lane lines, zebra crossings, road signs, etc. And the problem of poor accuracy in two-dimensional scenes is supplemented.

[Key words] visual SLAM; GPS; fusion positioning; intelligent vehicle positioning; lightweight network

1 基于道路场景建模的自动驾驶车辆定位

自从 SLAM 方法问世以后, 自动驾驶车辆的定位方法就分成了两大类。一类是以激光和视觉相机进行的实时定位与建图, 即 SLAM。另一类则是根据已经建立好的地图来进行定位的道路场景建模定位。

实时定位与建图是根据建图时所使用的传感器不同, 可再分为视觉 SLAM^[1] 和激光 SLAM^[2]。其中, 视觉 SLAM 方法主要流程可表述为: 前端、视觉里程计、后端、回环检测等。在小型家用扫地机器人等应用场景中已取得较好应用成效, 但是在户外无人驾驶领域的应用却仍亟待继续深入研究。

而基于道路场景建模的定位方法, 是先要建立道路的多维地图, 再与行驶过程中的地图进行匹配, 这种定位方法和平常在看地图寻找目的地时具有相同的逻辑过程。所以, 当车辆使用这种方法进行定

位时, 对道路的建模和表征, 每个节点所包含的信息的准确性, 以及相应的信息维度, 就是其获得定位精确性的根本和基础。

一般来说, 为了进行更加精确的定位, 在进行道路场景建模时, 就会使用精度较高的传感器。例如 GPS 和高精度惯性导航传感器、激光雷达、高清摄像头等等。在进行道路多维地图构建时, 建图车辆以某一固定的速度在地图标识路段行驶, 将采集多种数据信息, 例如 GPS 信号、惯性导航传感器信号、视觉图像等等。而在采集了多重传感器的数据后, 通过传感器标定, 数据融合, 将采集到的信息进行关联^[3-5]。

在道路场景建模完成后, 先通过普通的 GPS 进行定位, 并在所建立的模型中得到一定范围内的采集点, 接着对采集车采集到的各种信息, 诸如视觉图像、点云信息、惯导数据等与事先建立好的多维地图进行匹配。根据采集车选用传感器的不同, 道路多

作者简介: 种玉祥(1995-), 男, 硕士研究生, 主要研究方向: 视觉 SLAM。

通讯作者: 种玉祥 Email: 854967515@qq.com

收稿日期: 2020-09-27

哈尔滨工业大学主办 ◆ 学术研究与应用

维地图通常可分为2类,即:基于视觉摄像头的多维地图构建;基于高精度激光雷达的多维地图构建。

和采用激光雷达相比,运用相机进行地图的构建和定位可以大大降低制图与定位的成本,所以,近年来,多种视觉地图构建方法陆续涌现出来。本文也提出了一种视觉地图构建方式,该方法包含诸如节点位姿、二维场景特征、道路标志特征等节点要素,如图1所示。

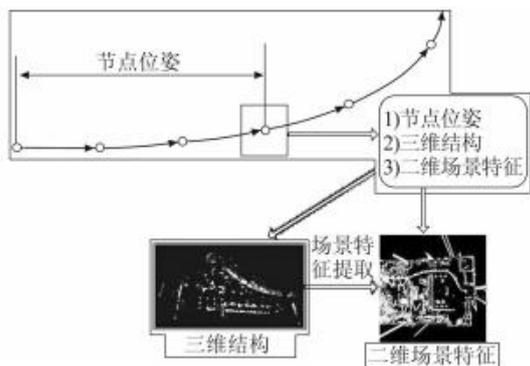


图1 道路场景表征模型的节点要素

Fig. 1 Node elements of road scene representation model

其中,道路二维投影特征由文中提出的区域点云加权投影方法,由优化后的点云图中运算获取。节点位姿则可以通过传统的视觉里程计算法等获取。道路标识特征由轻量级神经网络识别得到。

本文提出的节点要素获取方法将在下文进行阐释与分析。

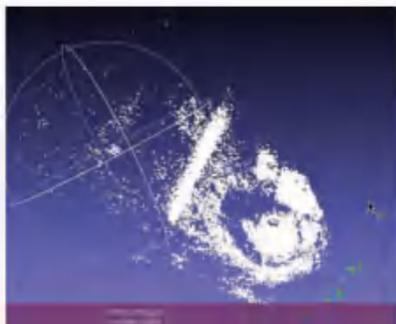
2 道路投影特征

不管是基于视觉、还是激光所构建的多维地图,都是由大量的数据采集节点组成,在每个节点,都要采集视觉数据、GPS 信息、以及与 RTK 组合系统采集的高精度位置信息。为了使文中构建的视觉多维地图能够更加准确,本文在每个节点选择存储的信息分别为:节点位姿、道路投影特征、道路标志特征、GPS 信号等。

2.1 点云预处理

本方法要对三维点云进行处理,用来形成二维特征。车载系统对计算性能存储容量有较高的要求。不论是出于降低所占用的存储资源,还是去除噪声,降低运算负荷等目的,都要对点云进行预先处理。在预处理时,为保证点云的质量,需要对所选定的点云的区域范围进行选择。首先考虑的目标是要选择的区域,点云要尽可能地稠密,同时,还要满足成像清晰,误差小等要求。考虑到是用作建图来使用,所以只需要保留道路信息,对于采集时得到的车

辆和行人等动态目标要采用滤波器进行剔除,防止这些因素影响构建好的地图使用时的配准率。BAL 数据集在使用 Ceres BA 优化后的可视化点云图,如图2所示。



(a) 原始的点云信息

(a) Original point cloud information



(b) 预处理后的点云信息

(b) Preprocessed point cloud information

图2 点云预处理

Fig. 2 Point cloud preprocessing

2.2 区域点云加权投影方法

传统的经典配准算法,主要有 ICP。虽然定位精度高,但是运行速度慢,消耗的时间过长。当用于车辆自动驾驶时,研究发现最终准确度达到了较好效果,其运行速度却过慢,因而不能满足智能车辆行驶时对即时性的要求。基于此,本文提出了一种视觉点云区域加权投影方法,通过这种方法,能够将研究中事先进行过预处理的点云转换成二维场景特征,继而采用目前一些运行速度较快,稳定性较高的视觉特征提取算法进行特征匹配与特征提取,经仿真验证可知,这对提升点云的配准效率有较大的帮助。

研究中假设:安装在车辆上的摄像头,其 Z 轴与地平面是相互垂直的。并且,车辆一般都是行驶在结构化道路上,而摄像机镜头的坐标在垂直方向上,即 Z 轴上的微小变化可以近似地认为是接近于零的。因此,在对点云事先进行处理后,仅需将之前重建后的点云图投影到地表平面即可。先将处理后

的点云, 划分为一个个长方形区域, 设为 $2A_1$ 行和 $2A_2$ 列。设视觉点云 P 中的一点为 $p_i(x, y, z)$, 其中 x 的取值范围为 $(-A_1, A_1)m$, y 的取值范围为 $(-A_2, A_2)m$, 将视觉点云投影在地面平面上。因此二维场景图像的大小为 $(2A_1 \times 2A_2)$ 像素, 而图像中的每个像素值 $J(i, j)$ 由公式(1)来求得:

$$J(i, j) = \begin{cases} 0, & \sum p_{i,j} < Q_{i,j}, \\ 255, & \sum p_{i,j} \geq Q_{i,j}. \end{cases} \quad (1)$$

其中, $\sum p_{ij}$ 是第 i 行第 j 列长方形区域中所有点云的数量之和, $Q_{i,j}$ 是该长方形区域的投影权重, 权重的大小受 Z 轴一定距离内点云分布的影响。具体如公式(2)所示:

$$Q_{i,j} = \frac{\sum p_{-A_1, -A_2} + \dots + \sum p_{A_1, A_2}}{2A_1 \times 2A_2}. \quad (2)$$

通过这种方法, 就能获得视觉点云转换成的二维场景图像, 图像大小为 $(2A_1, 2A_2)$ 像素。本文提出的视觉点云加权投影方法示意图参见图 3。

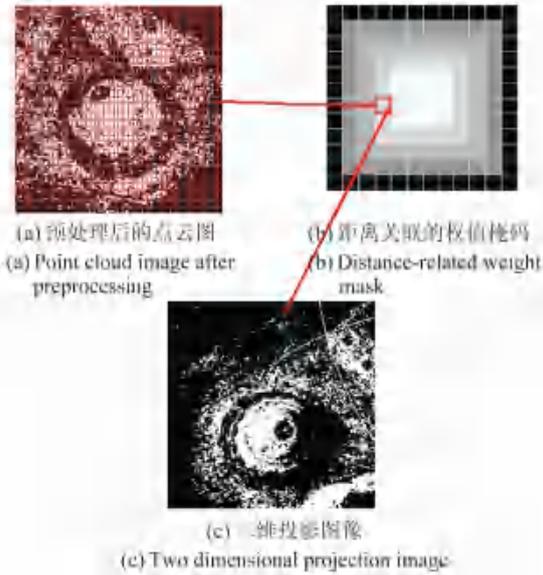
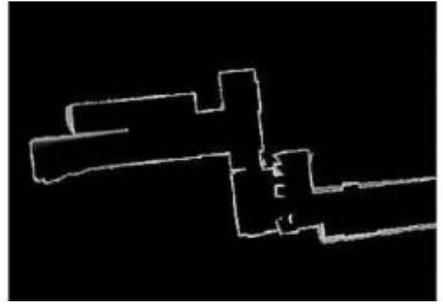


图 3 距离关联的视觉点云加权投影方法示意图

Fig. 3 Schematic diagram of distance-related visual point cloud weighted projection method

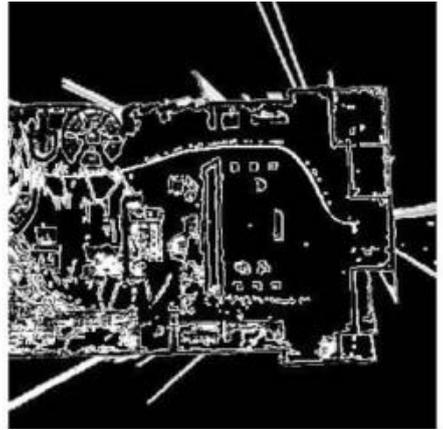
此外, 在将点云投影到地表平面上时, 难免会丢失垂直方向上的信息。因此, 在对点云进行投影时, 计算研究过程中划分的小区域内点云的平均高度, 以此来保留 Z 轴上的高度信息。

在校区教学楼附近运行时生成的部分二维场景图像如图 4 所示。



(a) 二维场景图像一

(a) Two-dimensional scene image I



(b) 二维场景图像二

(b) Two-dimensional scene image II

图 4 二维场景特征图像

Fig. 4 Two-dimensional scene feature image

2.3 ORB 特征提取算子

在获取了由视觉信息转换后的二维场景地图后, 如何运用一种图像特征提取方法, 才能在不损失准确度的前提下尽可能地提高运算速度则是本次研究中在构建视觉多维地图时需要考虑的重点问题。目前, SIFT^[6] 与 SURF^[7] 是热门流行的特征提取算子, 具有良好的性能, 识别准确, 但是这两种算法在运行速度上都有些不足, 特别是在行驶的车辆上使用这一不足就体现得更加明显, 车辆在行驶时对定位和匹配的实时性就提出了很高的要求, 因此, 运行速度慢就成了很大的缺点。综上, 为了解决前述问题, Rublee 等人^[8] 提出了 ORB 特征检测算法, 和 SIFT、SURF 相比, 该算法具有更加优良的性能。ORB 算法的特征提取方法是检测相邻域中像素的灰度以及角点, 因此该方法在保证检测效率前提下, 对速度有较大的提升。

ORB 算子特征匹配示意图如图 5 所示。ORB 算子在图片发生拉伸等情形时, 仍能对所获取的图像进行较准确的识别与匹配, 能够更好地满足智能驾驶中二维场景的匹配与定位需求。



图5 特征提取示意图

Fig. 5 Feature extraction diagram

3 车道标志信息获取

本小节提出了一种轻量级的神经网络,对道路特征标志,如车道线、斑马线、路牌等信息进行识别和记录,和前文获取的视觉点云,二维结构等信息相结合,使建模时采集的每一个节点都具有不可替代性,用以增加使用地图定位匹配时的准确度。

由于内存和计算资源有限,现实中在车辆上部署深度神经网络较为困难。基于此,本文设计了一种全新的神经网络基本单元 Shadow 模块,并搭建出轻量级神经网络架构 ShadowNet。在一个训练好的深度神经网络中,通常会包含丰富甚至冗余的特征图,以保证对输入数据有全面的理解。但是,并非所有特征图的获取都要用到卷积操作,也可通过线性运算与恒等映射的操作来生成。

鉴于主流神经网络计算出的中间特征图中存在大量的冗余,为了减少所需的资源,Shadow 的主要作用是生成其中的卷积核。首先,输入数据 $X \in R^{c \times h \times w}$,这里 c 是输入通道数, w 和 h 是输入数据的宽度与高度,所以,生成特征图的卷积层可用如下运算来表示:

$$Y = X * f + b, \quad (3)$$

其中,“ $*$ ”运算表示卷积运算; b 表示偏差项; $Y \in R^{\hat{h} \times \hat{w} \times n}$ 表示输出特征图,具有 n 个通道; $f \in R^{k \times k \times c \times n}$ 为该层中的一个卷积核。这里, \hat{h} 和 \hat{w} 分别表示输出数据的高度和宽度, $k \times k$ 表示卷积核 f 的内核大小。

利用式(3),就可以通过输出特征图的尺寸,确定研究中要优化的参数数量(f 和 b)的值。卷积层的输出特征图一般都有很多冗余特征,而且其中一些还具有较高的相似度,因此可通过运算公式进行恒等映射,将特征图进行去冗余处理。此处拟用到的数学公式可写为:

$$\hat{Y} = X * \hat{f}, \quad (4)$$

其中, $\hat{f} \in R^{c \times k \times k \times m}$ 表示使用的卷积核,且 $m < n$,为简单起见,这里省略了偏差项。超参数与普通卷积中的超参数相同,如此可使得输出特征图的空间大小保持一致。为了进一步获得所需的 n 个特征图,继续对 \hat{Y} 中的每个原始特征应用一系列的线性运算,以生成 s 个 Shadow 特征图,即:

$$y_{i,j} = \Phi_{i,j}(y'_i), \forall i = 1, \dots, m, j = 1, \dots, s. \quad (5)$$

其中, y'_i 表示 \hat{Y}_i 中第 i 个原始特征图, $\Phi_{i,j}$ 表示第 j 个线性运算,用于生成第 j 个 Shadow 特征图。

为了提升模型特征表达能力,避免设计的网络受到低计算负载的约束而导致性能降低,本算法将通过线性运算的 s 个 Shadow 特征图与在去冗余后输出的特征图进行加权叠加,将最后的结果 $Y = [y_{11}, y_{12}, \dots, y_{ms}]$ 用来做 Shadow 模块的输出数据。具体操作如图 6 所示。

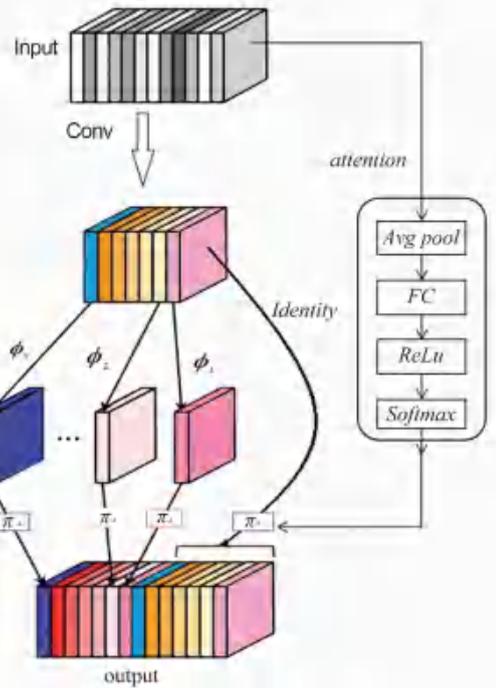


图6 Shadow 模块

Fig. 6 Shadow module

通过上述 Shadow 的模块,构建出专门为移动设备中神经网络设计的 Shadow bottleneck。Shadow bottleneck 主要由 2 个堆叠的 Shadow 模块组成。第一个 Shadow 模块用作扩展层,增加了通道数。第二个 Shadow 模块减少通道数,以与 shortcut 路径匹配。通过使用 shortcut 连接与这两个 Shadow 模块的输入、输出相连接,能够很好地减少计算资源的消耗,降低模型的过拟合,有效缓解梯度消失或梯度爆炸。

在道路实际模拟时采集的道路特征信息如图 7 所示。



(a) 特征检测结果一

(a) Feature detection result I



(b) 特征检测结果二

(b) Feature detection result II

图7 道路标志特征检测结果

Fig. 7 Road sign feature detection result

4 基于视觉的多维地图构建方法综述

本文提出的基于视觉的多维地图构建方法总体可以归纳为如下步骤:

(1) 运用车载单目摄像机、车载 GPS 等结合惯性导航传感器,对需要建立视觉多维地图的路段进行数据采集,并做三维重建,在获取其他信息的同时,也得到了点云信息。

(2) 对生成点云进行加权二维投影,对处理后的二维场景采用 ORB 算子进行特征提取,用以进行节点匹配。

(3) 由于点云转二维后场景清晰度不足,本文在这里采用的是构建轻量级神经网络的方法,对道路标志特征如车道线、斑马线、道路交通标志牌等进行识别与记录,与 GPS 定位,以及(2)中提及的道路二维特征进行同步匹配,使每个节点更加具有独特性,增加匹配准确率。

(4) 本文的建图方法中,每个节点都存储了点云信息、二维信息、道路标志特征信息。在定位时可先利用 GPS 等进行粗定位,再借由二维场景特征和道路特征信息进行精确定位,最后运用点云和轨迹信息进行空间位姿的精确定位。

5 实验结果分析

本文试验采用上海工程技术大学参与改进的无人驾驶试验采集车进行数据的获取。实验的测试路段选择在上海工程技术大学内的教学楼附近,试验

路段的俯瞰图及试验路段道路场景如图 8 所示。由图 8 可知,试验路段完全满足机动车行驶要求,且道路行车路线和路面都具有—般性,故认为实验结果可作为最终结论。本次实验的实验线路总长度约为 900 m。



(a) 试验路线图

(a) Test roadmap



(b) 道路场景图

(b) Road scene graph

图8 试验路段场景图

Fig. 8 Scene map of test section

根据本文提出方法进行地图构建时,先使用车载相机对道路信息进行采集,运用视觉 SLAM 技术生成点云,同时对点云进行处理,获得二维投影信息;在获取了点云、位姿和二维投影信息后,运用轻量级神经网络对道路标志特征如:车道线、路灯、告示牌等进行道路标志特征提取。在构建的众多信息采集点中,随机选择 1~2 个作为多维地图的起点,并进行特征匹配。通过视觉 SLAM 方法获得视觉轨迹,最后结合视觉轨迹和地理轨迹的相对位置关系,生成视觉地图。

验证环节主要对视觉轨迹产生的累积误差进行分析。针对多维信息构建出的视觉多维地图,在分析误差时同样要进行多维分析。本文以每 10 m 为一个单元,选取了长度为 100 m 的实验路段,用以进行累积误差的分析。测试结果如图 9 所示。由图 9 可知,当测试路段长度在 10~90 m 之间时,累积误差较低,仅在 0.5 m 以下;当测量长度为 100 m 时,累积误差也能控制在 0.5 m 左右。

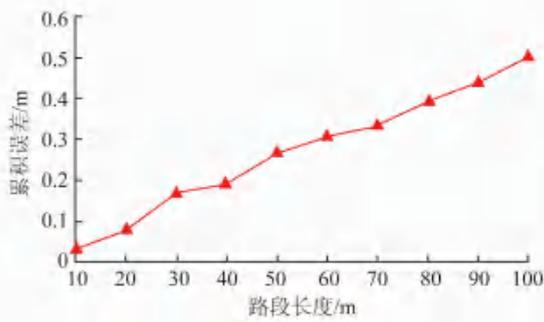


图9 视觉轨迹图与累积误差图

Fig. 9 Visual trajectory graph and cumulative error graph

6 结束语

本文提出了一种基于视觉的多维地图构建方法。首先利用视觉 SLAM 原理获取点云信息,再使用视觉点云的区域加权三维投影方法,生成二维场景表征图像,使用 ORB 特征提取算子对生成的二维表征图像进行特征提取。对于生成的二维场景表征图像准确率低的问题,构建了新的轻量级神经网络,识别道路标志特征,作为补充,增强每个节点的独特性,提高视觉配准的准确率。同时,也记录 GPS 等信息,在定位时先通过 GPS 信号进行粗定位,提高了定位的速度。经试验验证,地图制图误差在合理

范围内。

参考文献

- [1] MAHON I, WILLIAMS S B, PIZARRO O, et al. Efficient view-based SLAM using visual loop closures[J]. IEEE Transactions on Robotics, 2008, 24(5):1002-1014.
- [2] TANG Jian, CHEN Yuwei, NIU Xiaoji, et al. LiDAR scan matching aided inertial navigation system in GNSS-denied environments[J]. Sensors, 2015, 15(7):16710-16728.
- [3] ZHANG Z. A flexible new technique for camera calibration[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(11):1330-1334.
- [4] HWANGBO M, KIM J S, KANADE T. IMU self-calibration using factorization[J]. IEEE Transactions on Robotics, 2013, 29(2):493-507.
- [5] Bouquet J Y. Camera calibration toolbox for Matlab[EB/OL]. [2019-05-05]. http://www.vision.caltech.edu/bouquetj/calib_doc/.
- [6] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2):91-110.
- [7] BAY H, ESS A, TUYTELAARS T, et al. Speeded-Up Robust Features (SURF)[J]. Computer Vision & Image Understanding, 2008, 110(3):346-359.
- [8] RUBLEE E, RABAUD V, KONOLIGE K, et al. ORB: An efficient alternative to SIFT or SURF [C]// International Conference on Computer Vision. Barcelona, Spain; IEEE, 2012: 2564-2571.
- [3] JIN R, WANG G. Simple, fast, and scalable reachability Oracle [J]. Very Large Data Bases, 2014, 6(14): 1978-1989.
- [4] JIN Ruoming, XIANG Yang, RUAN Ning, et al. Efficiently answering reachability queries on very large directed graphs[C]// Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data. Vancouver, BC, Canada: ACM, 2008: 595-608.
- [5] YILDIRIM H, CHAOJI V, ZAKI M J. GRAIL: A scalable index for reachability queries in very large graphs [J]. The VLDB Journal, 2012, 21(4): 509-534.
- [6] ZHANG Tianming, GAO Yunjun, LI Congzheng, et al. Distributed reachability queries on massive graphs[M]//LI G, YANG J, GAMA J, et al. Database Systems for Advanced Applications. DASFAA 2019. Lecture Notes in Computer Science. Cham; Springer, 2019, 11448: 406-410.
- [7] SENGUPTA N, BAGCHI A, RAMANATH M, et al. ARROW: Approximating reachability using random walks over Web-scale graphs [C]// International Conference on Data Engineering. Macao, China; dblp, 2019: 470-481.
- [8] BENDER M A, FINEMAN J T, GILBERT S, et al. A new approach to incremental topological ordering[C]// Symposium on Discrete Algorithms. Austin, Texas; dblp, 2009: 1108-1115.
- [9] JAGADISH H V. A compression technique to materialize transitive closure [J]. ACM Transactions on Database Systems, 1990, 15(4): 558-598.
- [10] CHEN Y, CHEN Y. An Efficient algorithm for answering graph reachability queries [C]// IEEE 24th International Conference on Data Engineering. Paris, France; IEEE, 2008: 893-902.
- [11] YILDIRIM H, CHAOJI V, ZAKI M J. Grail: Scalable reachability index for large graphs [J]. Proceedings of the VLDB Endowment, 2010, 3(12): 276-284.
- [12] VELOSO R R, CERF L, Jr MEIRA W, et al. Reachability queries in very large graphs: A fast refined online search approach [C]//17th International Conference on Extending Database Technology. Athens, Greece; dblp, 2014: 511-522.
- [13] TANG Xian, CHEN Ziyang, LI Kai, et al. Efficient computation of the transitive closure size [J]. Clust. Comput., 2019, 22 (Supplement): 6517-6527.
- [14] JIN R, RUAN R, DEY S, et al. SCARAB: Scaling reachability computation on large graphs [C]// Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data. Scottsdale: ACM, 2012: 169-180.
- [15] CHA M, HADDADI H, BENEVENUTO F, et al. Measuring user influence in Twitter: The million follower fallacy [C]// Proceedings of the Fourth International Conference on Weblogs and Social Media (ICWSM 2010). Washington, DC, USA; dblp, 2010: 10-17.
- [16] Van SCHAICK S J, De MOOR O. A memory efficient reachability data structure through bit vector compression [C]// Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data. Athens, Greece; ACM, 2011: 913-924.

(上接第28页)