|  |  |
| --- | --- |
| **学号** | **2112233062** |

**广 州 大 学**

**硕士研究生学位论文开题报告**

姓 名: 何智鹏

学 号： 2112233062

专业\领域： 网络与信息安全

论文题目： 基于地标技术的最短距离查询

与三角形计数算法研究

指导教师： 欧阳典

是否保密： 否

保密等级： 无

2023年 4 月20 日填

**填表说明：**

1. 表格要求统一以“小4号宋体”字体填写，双面打印，表格的格式和页码次序不能擅自改变，如需附页一律加在本表格后面，并注明前后接连地方；
2. 所有栏目须填写清楚，不留空白，所有审核意见须填写具体、有针对性；
3. 凡没有建立班组或系教研室的，基层组织意见或系教研室意见一律由学院主管院长负责填写：
4. 签名必须由相应责任人签名，不得打印代替；
5. 须盖学院公章的地方不能不盖；

6、本表一式三份，学院审核后，研究生本人留一份，学院留存一份，报研究生院一份。

|  |  |
| --- | --- |
| **学位论文题目** | 基于地标技术的最短距离查询与三角形计数算法研究 |
| **经费来源** | 大规模道路网络上基于索引的最短路径问题研究 |
| **本研究题目是否属导师课题的一部分？如果是的话，属哪一部分？**  是的，属于“2023A1515110592 《大规模道路网络上基于索引的最短路径问题研究》”这一部分。 | |
| **学 位 论 文 的 类 型**（在相应栏目划√）**：**  **☑理论研究 □应用基础 □用于生产 □其他** | |
| **一、立题依据**（包括国内外研究现状和发展趋势、文献综述、研究目的、意义等）**：**  **1.1 研究背景**  随着大数据和人工智能等新兴技术的发展，图数据的应用也日益受到关注。在现代社会，互联网和各种数字化终端普及，导致全球每年产生的数据规模达到了ZB级别。这些数据不仅记录着世界上发生的事件，还反映着事物之间错综复杂的关系。因此，如何快速获取所需信息成为一项迫切需要解决的问题。  图作为一种重要的数据结构，在大数据和人工智能的推动下迎来了长足的发展和广泛的应用。根据IT研究与顾问咨询公司Gartner的报告，图被认为是十大数据与分析技术趋势之一，并且在数据分析和数据挖掘中扮演着重要角色。同时，2019年-2023年全球图处理与图数据库应用呈现出每年100%的增长速度，进一步凸显了图数据在现代数据分析中的重要性。  近年来，知识图谱作为一种特殊类型的图数据备受学术界和工业界的关注。越来越多的国内外企业开始利用大规模知识图谱解决商业、经济、生命科学等多个领域的实际问题。例如，全球领先的电商平台利用知识图谱技术构建用户画像，实现了个性化推荐和精准营销，极大提高了用户购物体验和交易转化率。另外，医疗保健领域也积极应用知识图谱技术，通过整合医疗数据和研究成果，加速了疾病诊断和治疗方案的优化。此外，金融行业借助知识图谱技术构建风险控制系统，实现了对交易欺诈和信用风险的有效监测和预防。这些实际案例充分展示了知识图谱在不同领域的广泛应用和重要价值。  在学术界和业界的共同努力下，一系列图数据处理技术不断涌现，并取得了显著的进展。可达性查询[1]、最短路径查询[2]、简单路径枚举[3]、环路检测[4]等技术在同构图上已经有了较为成熟的应用。然而，随着数据量和数据关系的增加，现有的图数据处理技术仍然存在着一定的局限性，尤其是在数据量在数百万、数亿级别的数据查询和分析方面，需要进一步的研究和探索。  **1.2 研究现状**  由于最短路径查询和三角形计数在图数据分析中有重要作用，以下对现阶段的相关工作进行总结。  **最短距离查询方面：**  **（1）基于搜索的方法**  为了准确计算最短路径距离，一种简单的方法是在加权图中使用Dijkstra算法，或者在无权图中使用广度优先搜索（BFS），从一个源点到目标点进行搜索[5]。为了提高搜索效率，可以采用双向搜索方案，即同时从源点和目标点进行搜索[6]。后来，Goldberg等人将双向搜索与A算法结合[7]，以进一步提高搜索性能。在这种方法中，他们预先基于地标计算标签来估算最小距离，并将此估算用于双向A搜索，从而高效地计算最短路径距离。然而，这种方法主要适用于道路网络，并不适合于大多数其他复杂网络[8]。  **（2）基于标签的方法**  为了有效地解答图中精确的最短路径距离查询，研究者们已经成功开发了基于标签的方法[8-10]。这些方法大多基于2跳覆盖的概念来构建标签[11]。已经证明，计算最小的2跳覆盖标签是NP-Hard问题[10, 12]。在一项研究中[8]，作者提出了一种hub-based的标签算法（HL），该算法通过处理收缩层次结构来构建hub标签，它是已知最快的用于道路网络距离查询的算法之一。然而，这种方法并不适用于复杂网络。因此，同一作者在另一项研究中提出了用于复杂网络的分层hub标签算法（HHL）[9]。在这项工作中，采用自上而下的方法维护每个顶点的最短路径树，以指示每个顶点的所有未覆盖的最短路径。由于非常高的存储和计算需求，这种方法也不适合处理大型图。此外，Jin等人[13]提出了一种称为高速公路中心标签（HCL）的方法，该方法利用图的高速公路结构。这种方法旨在找到一个可以帮助进行最优距离标记的生成树，并使用该生成树作为高速公路来计算基于高速公路的2跳标签，以快速计算距离。此后，Akiba等人[10]提出了修剪地标标签（PLL）方法，该方法通过从每个顶点进行修剪的广度优先搜索（BFS）预先计算一个距离感知的2跳覆盖索引。其想法是修剪一些顶点，这些顶点的距离信息可以通过通过以前的BFS构造的部分可用的2跳索引来获得。这项工作由于减少了搜索空间，因此在百万级网络上实现了较低的构建成本和较小的索引大小。研究表明，PLL在发布时超越了当时可用的其他最先进的方法，包括HHL[9]、HCL[13]和TEDI[14]。然而，由于对标签构建的极高内存需求，PLL仍然不适合为十亿级网络构建2跳覆盖索引。  **（3）基于预计算聚类距离的方法**  为了解决传统算法在大规模数据面前效率不高的问题，JENS MAUE[15]等人提出了一种使用PCD(预计算聚类距离)进行目标导向的最短路径查询算法，该算法的核心思想是先将整个网络划分为多个聚类，然后预先计算并存储聚类内部和聚类之间的最短路径距离。首先，基于网络结构特点进行聚类，将网络节点分组以形成多个簇；其次，计算簇内部的最短路径并存储结果；接着，计算并存储不同簇之间的代表节点的最短路径。这样，在进行实际的最短路径查询时，可以直接利用这些预存的距离信息，通过组合不同簇之间的路径来迅速得出整个网络的最短路径。使用PCD算法，可以大幅度降低实时计算的需求，显著提高路径查询的响应速度。这对于需要快速处理大量路径查询的应用（如在线地图服务、网络数据分析等）尤为重要。  **精确三角形计数方面：**  **（1）基于矩阵乘法的方法**  这种方法将图的连接关系表示为一个矩阵（邻接矩阵）。通过对这个矩阵进行特定的数学运算（计算其三次幂并考察特定位置的值），可以直接得到图中三角形的精确数量。这种方法的理论基础很清晰，能够保证结果完全准确。但它的主要缺点是进行这些矩阵运算需要非常大的计算时间和内存空间，特别是当图的节点很多时，会变得不切实际。  **（2）基于节点/边迭代与邻域相交的方法**  这是目前处理大规模图（特别是稀疏图）最主流且实用的精确计数方法。它通过系统性地遍历图中的节点或边，并检查它们的邻居关系来确保找到所有三角形。例如，对每条边 (u, v)，计算其端点邻居集合 N(u) 和 N(v) 的交集，交集中的每个节点 w 都构成一个精确的三角形 (u, v, w)。通过结合“度排序”（将边从低度节点指向高度节点）等关键优化，可以减少冗余计算。  **1.3 研究意义**  **（1）加速大图最短路径查询算法设计的价值**  回答精确的最短路径距离查询。毫无疑问，这是图论中的一项基本任务，它有着广泛的应用。例如，在社交网络上，两个用户之间的距离被认为表示亲密程度，并用于社交敏感搜索，以帮助用户找到更多相关的用户或内容，或者是分析有影响力的人和社区。在网络图上，网页之间的距离是相关性的指标之一，并用于上下文感知搜索，以对与当前访问的网页更相关的网页给予更高的排名。距离查询的其他应用包括对链接数据的top-k关键字查询、发现代谢网络中化合物之间的最佳途径以及管理计算机网络中的资源。  **（2）现有算法的局限性**  当涉及到大型图形时，我们可以使用宽度优先搜索（BFS）或Dijkstra算法计算每个查询的距离。但是这些算法需要一定的计算时间，特别是对于社交敏感搜索或上下文感知搜索等需要实时交互的应用程序，它们需要多对顶点之间的距离来对每个搜索查询的项目进行排序。因此，距离查询的回答速度应该快得多，比如说，微秒。另一种极端方法是预先计算所有顶点对之间的距离，并将其存储在索引中。虽然我们可以立即回答距离查询，但这种方法也是不可接受的，因为预处理时间和索引大小都是平方级的，而且非常大。  **（1）加速大图三角形计数算法设计的价值**  三角形计数算法的研究具有重要意义。首先，三角形是网络中最基本的非平凡结构单元，其数量和分布（如聚类系数）是衡量网络紧密程度、社群结构和传递性的核心指标。其次，高效计数三角形有助于深入理解社交网络、生物网络、金融网络等的拓扑特性、演化规律和功能模块。此外，它也是图数据挖掘、大规模图算法设计与优化（如图数据库查询、并行/流处理）的一个基础且具代表性的问题，推动着相关计算技术的发展。  因此，开发一种支持广泛应用于大量图形数据的高效最短路径距离查询算法和三角计数对于图数据分析具有显著的重要性。  **【参考文献】**   1. Su J, Zhu Q, Wei H, et al. Reachability querying: Can it be even faster?[J]. IEEE Transactions on Knowledge and Data Engineering, 2016, 29(3): 683-697. 2. Ma S, Feng K, Li J, et al. Proxies for shortest path and distance queries[J]. IEEE Transactions on Knowledge and Data Engineering, 2016, 28(7): 1835-1850. 3. Peng Y, Zhang Y, Lin X, et al. Hop-constrained s-t Simple Path Enumeration: Towards Bridging Theory and Practice[J]. Proc. VLDB Endow., 2019, 13(4): 463-476. 4. Qiu X, Cen W, Qian Z, et al. Real-time constrained cycle detection in large dynamic graphs[J]. Proceedings of the VLDB Endowment, 2018, 11(12): 1876-1888. 5. Robert Endre Tarjan. 1983. Data structures and network algorithms. Vol. 44. Siam. 6. Ira Pohl. 1971. Bi-derectional search. Machine intelligence 6 (1971), 127–140. 7. Andrew V Goldberg and Chris Harrelson. 2005. Computing the shortest path: A search meets graph theory. In SODA. 156–165. 8. Ittai Abraham, Daniel Delling, Andrew V Goldberg, and Renato F Werneck. 2011. A hub-based labeling algorithm for shortest paths in road networks. In SEA. 230–241. 9. Ittai Abraham, Daniel Delling, Andrew V Goldberg, and Renato F Werneck. 2012. Hierarchical hub labelings for shortest paths. In ESA. 24–35. 10. Yen J Y. Finding the k shortest loopless paths in a network[J]. management Science, 1971, 17(11): 712-716. 11. Rizzi R, Sacomoto G, Sagot M F. Efficiently listing bounded length st-paths[C]//Combinatorial Algorithms: 25th International Workshop, IWOCA 2014, Duluth, MN, USA, October 15-17, 2014, Revised Selected Papers. Cham: Springer International Publishing, 2015: 318-329. 12. Grossi R, Marino A, Versari L. Efficient algorithms for listing k disjoint st-paths in graphs[C]//LATIN 2018: Theoretical Informatics: 13th Latin American Symposium, Buenos Aires, Argentina, April 16-19, 2018, Proceedings 13. Springer International Publishing, 2018: 544-557. 13. Peng Y, Zhang Y, Lin X, et al. Hop-constrained s-t Simple Path Enumeration: Towards Bridging Theory and Practice[J]. Proc. VLDB Endow., 2019, 13(4): 463-476. 14. Sun S, Chen Y, He B, et al. Pathenum: towards real-time hop-constrained s-t path enumeration[C]//Proceedings of the 2021 International Conference on Management of Data. 2021: 1758-1770. 15. Tarjan R. Enumeration of the elementary circuits of a directed graph[J]. SIAM Journal on Computing, 1973, 2(3): 211-216.s 16. Al Hasan M, Dave V S. Triangle counting in large networks: a review[J]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2018, 8(2): e1226. 17. Azad A, Buluç A, Gilbert J. Parallel triangle counting and enumeration using matrix algebra[C]//2015 IEEE International Parallel and Distributed Processing Symposium Workshop. IEEE, 2015: 804-811. 18. Green O, Yalamanchili P, Munguía L M. Fast triangle counting on the GPU[C]//2014 4th Workshop on Irregular Applications: Architectures and Algorithms (IA^ 3). IEEE, 2014: 1-8. | |

|  |
| --- |
| 二、研究内容（说明课题的具体内容，独创及新颖之处，重点解决的问题，预期的研究成果）：  **2.1 研究内容**  本研究旨在探索一种新的图中重要顶点选取策略，即选取landmark，以替代传统方法中直接选择度数最高的顶点，从而能加速最短距离查询和三角形计数算法。该策略的主要思路包括：首先，在初始化阶段选择前K个顶点作为基准的landmark；其次，对于剩余的顶点，将它们划分到离其最近的landmark所在的区域中；最后，在每个区域中选择具有次高度数中心性的节点作为新的landmark。这一策略考虑了顶点之间的结构关系，更符合实际情况，并有望提高选取的landmark的代表性和算法的性能。  本课题的主要研究内容是设计出高效的查询算法和三角形计数算法以解决上述两类问题。设计出的算法应当适应不同类型的大规模图数据网络，如社交网络、信任网络、知识图谱等多种类型，在不同量级的图数据网络中均表现出其相比于以前算法的优势。  **2.2 关键问题**  现有最先进的算法是以landmark labelling-based，他们通常需要输入一个k，这个k代表landmark的个数，然后直接选择图中前k个degree值最高的顶点来作landmark，没有利用图结构的关键特征，忽略了顶点之间微妙的关系和相互作用，我们认为还有优化空间。  为了尝试不同的策略，我们最初考虑了使用前k个最高介数中心性的节点作为landmark。然而，尽管在标签大小方面有些提升，但在实际查询中并没有表现出显著的改善。更为重要的是，由于计算介数中心性在大图中的高耗时性，我们不得不舍弃这一方法。  通过对直接使用最高度数和最高介数中心性的节点作为landmark的策略进行分析，我们发现了两个主要问题：首先，这种方法选出的landmark与degree选出来的顶点存在重叠，且这种情况在大型网络中尤为突出。其次，选出的节点相对聚集，而我们希望它们能够分散开来，以更好地涵盖整个图。  在这个背景下，我们寻找了新的算法灵感，并受到了k-means算法的启发。我们提出了一种新颖的landmark选择策略：首先，我们将前k个节点作为基准landmark，将其它节点分配到离它们最近的landmark所在的区域。接下来，在每个区域中，我们再选择具有次高度数的节点作为新的landmark。这一过程旨在解决重叠和节点聚集的问题，提高查询效率。  值得注意的是，我们特别强调了在新landmark选择中只关注各自区域内的最高度数节点，而不考虑跨区域的节点。这个细节确保了新的landmark分散在整个图中，而不是集中在某个局部区域。  **2.3 独创及新颖之处**  （1）提出了一种创新的图中重要顶点（landmark）选取策略，相较于传统的最高度数选择方法，该策略更好地捕捉了图结构的重要特征，有效减少了Landmark的重叠率和聚集性。  （2）引入了区域划分的概念，通过将顶点划分到最近的Landmark所在的区域，使得Landmark在整个图上更加分散，减少了Landmark的聚集性，从而提高了查询效率。  （3）提出的Landmark选择策略和查询优化方法具有良好的可扩展性和通用性，适用于多种图数据库、社交网络等实际应用场景，在处理亿级数据集上依然表现出色，为相关领域的研究和应用提供了有力的支持。 |
| 三、科研方案设计（包括：研究方法、技术路线、理论分析、计算、实验方法和步骤及其可行性，可能出现的技术问题及解决方法）：  **3.1 研究方法**  本研究将采用深入分析、算法设计和实验验证相结合的方法。首先，通过对给定图的结构进行深入分析，了解其特点，为选取landmark提供理论基础。其次，设计基于区域划分和度数中心性的选取算法，以确保选取的landmark能够代表整个图的结构特征。最后，通过大量实验验证算法的有效性和性能，比较本方法与传统方法的优劣。  **3.2 算法设计**  算法设计包括三个主要步骤：初始化阶段、区域划分和新landmark的选择。在初始化阶段，选择前K个顶点作为初始的landmark。在区域划分阶段，将剩余的顶点划分到离它们最近的landmark所在的区域中。在新landmark的选择阶段，对每个区域选择具有最高度数的节点作为新的landmark。值得强调的是，我们特别注意了在新landmark选择中只关注各自区域内的最高度数节点，而不考虑跨区域的节点。这个细节确保了新的landmark分散在整个图中，而不是集中在某个局部区域。    图1. landmark的选择示意图    图2. 构建highway cover标签示意图    图3. 查询逻辑示意图    图3.1 查询同一个区域示意图 图3.2 查询不在一个区域顶点示意图  **3.3 实验验证**  实验验证分为前期验证和算法评估两部分。性能评估包括比较本方法与传统方法在不同规模的图上的性能差异，包括时间复杂度和空间复杂度。算法评估则通过对比选取的landmark与整个图的关键特征的一致性，评估算法的准确性和可靠性。  **（1）探索性实验**  测试随机批量生成的起点和重点，科学控制顶点命中同一区域的概率，让起点和重点尽量分散，初步体现基于区域查询思想的合理性。对比两种传统选取landmark策略的方法查询集下的算法效率。  **（2）有效性实验**  从理论上来讲，我们认为betweenness的定义更加契合landmark的初衷，通过实验验证猜想，发现使用Top-k betweenness值的顶点作为landmark提升效果并不明显，但在标签大小方面略微有提升。且随着图逐渐变大，标签大小越节约。但由于betweenness值在大图中的计算难度，我们放弃了单纯用betweenness值作为landmark的策略，在此思路上寻求新的选取landmark的策略。  在此基础上，我们需要探究为什么查询效率没有得到提高？于是我们对直接用Top-k degree作为landmark和直接用Top-k betweenness作为landmark选出来的顶点进行了分析，我们发现效果提升不明显的原因有两点：1. 发现各自方法选取出来的landmark是有重合的，且重合率较高。也就是说假若在大图中用betweenness作为landmark是有效果的，且在大图中提升效果会更显著。2. 选出来的点比较聚集，能尽量分散开最好。  **（3）索引成本**  对比不同landmark集合构建索引所需的时间以及索引占用的空间，分析索引开销。  **（4）可扩展性和算法查询性能实验**  在规模逐渐增大的路网上进行实验，对比不同规模路网上的查询效率，模拟批量查询，验证算法的并行处理表现和对批量查询的可扩展性。 |
| 四、基础及条件（包括：已经做过的有关研究工作，本单位或外单位可供使用的仪器设备和实验条件，已经获得或将要获得的经费等）：  **4.1 有关研究工作**  前期调研工作已完成。主要工作内容包括：搜集阅读中心性以及标签算法的相关论文，了解基于标签算法的最短路索引构建方式、时间复杂度、空间消耗等；针对基于标签实现的最短距离查询，了解不同算法之间的优劣等；了解最短距离查询的主要解决方案，为设计出高效的最短距离查询的大图算法提供前置知识。    图4.1 用Top-k betweenness值替代Top-k degree    图4.2 测试两种landmark顶点的重叠率  已完成的相关研究工作包括对图结构进行深入分析，了解图中顶点之间的关系，以及针对传统的landmark选取方法进行的调研和分析，发现其存在的局限性和不足之处。  在多个不同类型、不同规模的数据集上进行了大批量且精确值的最短距离查询的高效性实验，对比不同算法之间差异以及优劣，验证我们提出算法的高效性和扩展性。  **4.2 仪器设备和实验条件**  搭载Intel Xeon 2.10GHz CPU和256G运行内存的服务器，运行Ubuntu 20.04系统，实验相关代码均采用C++编写。 |

学 位 论 文 工 作 计 划

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 时 间 | 研 究 内 容 | | | | 预期效果 | | |
| 2023.09-2023.11 | 文献调研，阅读在正则路径查询的相关论文，归纳简单正则约束下的可达性分析以及路径枚举算法的不足之处 | | | | 对相关问题及算法研究进展有充分了解，针对不足之处进行完善与提升 | | |
| 2023.12-2024.02 | 进行理论分析与实验验证，分析算法的有效性和高效性 | | | | 初步完成算法设计 | | |
| 2024.03-2024.04 | 撰写开题报告，制作报告PPT | | | | 完成开题报告并通过审核 | | |
| 2024.05-2024.12 | 进行全面的实验分析，根据实验结果优化算法，进一步提升实验结果 | | | | 完成算法优化与实验验证 | | |
| 2025.01-2025.04 | 撰写论文 | | | | 提交毕业论文，通过答辩 | | |
| 论文工作起止日期 | 2023.09-2025.04 | | | 预计论文答辩日期 | | | 2025.05 |
| 科研经费开支计划 | | | | | | | |
| 低值设备仪器名称 | | 数 量 | 金 额 | 项 目 | | 金 额 | |
|  | |  |  | 试验材料费  试剂药品费  旅 差 费  资 料 费  其 它 | |  | |
| 合计金额：0元 | | | | | | | |
| **开题报告人签名：**  2024 年 4 月 29 日 | | | | | | | |

广州大学研究生学位论文开题报告审核表

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **研究生姓名** | | 何智鹏 | | | **导 师** | | 欧阳典 | |
| **专业\领域** | | 网络与信息安全 | | | **研究方向** | | 大规模网络信息的高效计算 | |
| **论文题目：**加速最短路距离查询的大图地标选取策略 | | | | | | | | |
| **开题报告时间：**2024年4月29日 | | | | | | **开题报告地点：**海丝知识中心B3-801 | | |
| 开题报告审核小组 | | | | | | | | |
| 组成 | 姓名 | | 职称 | 所在单位 | | | | 成员签名 |
| 组长 | 杨世宇 | | 教授 | 广州大学 | | | |  |
| 成员 | 卢璨 | | 副教授 | 广州大学 | | | |  |
| 张硕 | | 副教授 | 广州大学 | | | |  |
|  | |  |  | | | |  |
|  | |  |  | | | |  |
| 开题报告中提出的主要问题及回答的简要情况：  **杨世宇教授提问：**   1. 建议修改题目，以避免给人误解只是在做基于PCD的优化，限制太死了。   目前还完成论文攥写，所以还没能想好一个能更好概括全文中心思想的标题，对于提出的建议，我会认真考虑并修改题目，以确保其更具普适性和广泛性，避免被误解为仅限于基于PCD的优化。  **卢粲副教授提问：**   1. 对于你介绍的PCD，描述还不够清晰，需要进一步完善和详细解释。   PCD是预计算聚类距离的英文缩写，是一种预计算聚类距离进行目标导向的最短路径查询算法，该算法的核心思想是先将整个网络划分为多个聚类，然后预先计算并存储聚类内部和聚类之间的最短路径距离。确实需要更清晰地介绍其背后的原理和应用领域，以便读者更好地理解我们研究的背景和动机。  **张硕副教授提问：**   1. 有没有相关的论文成果？   目前有且仅有这一个研究内容，而且还处在优化算法设计阶段，可能还需要更多的实验数据来加强验证，所以还没有相关的论文成果，但我们计划在研究工作进一步深入后，撰写并提交相关的论文。 | | | | | | | | |
| 1. 那工作进度怎么样了？   目前，我们已完成了题目的初步设定和相关背景知识的调研工作，已经完成了算法的设计和实现，但尚有一些细节有待优化。工作进度已经完成起步的阶段了，我会保持紧张而有序的工作节奏，力争按计划推进研究工作。  记录人签名：  （不够写可加附页） 2024年 4 月 29日 | | | | | | | | |

|  |
| --- |
| 审核小组意见：（开题报告审核小组就论文选题意义及创新点，国内、外相关文献掌握程度和研究方法的可靠性，论文总体设计的科学性和可行性，存在的主要不足等提出意见或修改建议。）  该论文选题明确、背景新颖、算法设计有创新性。该生在论文选题之初，通过对国内外相关文献的深入研究，了解了研究领域的理论框架和现有方法，基本上了解论文选题所涉及的知识理论框架，并在开题报告中详细说明了该选题的研究意义、目的和主要研究内容及方法手段。其提出的替代传统的地标选择策略、区域划分算法以及通过区域哈希优化利用构建的标签在理论上可实施，且目前已经有了一定的实验结果。建议继续完成实验，同时实验效果的呈现需更加直观。  建议：1、通过。   1. 不通过，建议1个月内重新进行开题报告审核或终止学业。   组长签名：  （不够写可加附页） 年 月 日 |
| **学院审查意见:**  **主管领导签名： （公章）**  年 月 日 |

广州大学研究生院制表