

不确定网络条件可信近邻查询

郭长友^{*①②} 郑雪峰^① 高秀莲^②

^①(北京科技大学计算机与通信工程学院 北京 100083)

^②(德州学院 德州 253000)

摘要: 不确定因素在现实世界中普遍存在, 因此研究不确定网络条件可信近邻查询具有重要意义。该文给出一种新的解决方法。将不确定网络建模为不确定赋权图, 定义不确定图的样本图, 样本图指数, 基础网络, 可达路径长度及可达路径期望长度, 并给出基于不确定理论的高效不确定图条件可信近邻查询算法。将不确定网络上的近邻查询等价地转化为基础网络上的近邻查询问题。理论分析和实验结果表明所提可信近邻查询算法能够从非确定角度解决不确定网络环境下的近邻查询问题。

关键词: 不确定网络; 不确定图; 样本图指数; 可信距离; 基础网络

中图分类号: TP309.5

文献标识码: A

文章编号: 1009-5896(2016)04-0811-08

DOI: 10.11999/JEIT150748

Credible Nearest Neighbor Query in Uncertain Network

GUO Changyou^{①②} ZHENG Xuefeng^① GAO Xiulian^②

^①(School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China)

^②(Dezhou University, Dezhou 253000, China)

Abstract: Uncertain factors are a common phenomenon in the real world; therefore, it is very meaningful to study on the trusted neighbor query under uncertain network conditions. This paper puts forwards a new solution. The uncertain network is modeled as uncertain weighed graph, and these definitions of the uncertain graph are given, such as sample graph, sample graph index, base network, length of feasible path and expected length of feasible path. Based on these the high-efficiency credible neighbor query algorithm for uncertain graph is put forward under constraint conditions. This algorithm is transforms the issue of neighbor query in the uncertain network equivalently to the issue of neighbor query in the base network. The theoretic analysis and experimental results show that the credible neighbor query algorithm proposed in the paper can solve the neighbor query problem in the environment of the uncertain network from non-deterministic perspective.

Key words: Uncertain networks; Uncertain graph; Sample graph index; Credible distance; Basic networks

1 引言

现实世界中, 不确定因素普遍存在于各种现象中, 因此研究不确定网络的可信近邻查询具有重要意义。例如, 在云计算环境下, 数据中心的选址问题。云计算环境下, 虚拟机、虚拟集群的迁移、调度问题。Ad-hoc 网络的迁移、自组织问题, 传统网络模型无法刻画、表示、求解这些问题, 所以对不

确定网络的研究需要也越来越迫切。数据中心^[1]是为各种 IT 应用和服务提供的中心, 是集数据计算、交换和存储为一体的中心。数据中心主要由计算、网络、存储、灾备 4 个要素组成。网络是数据中心的基础设施, 包括交换机、连接线路、数据中心网络优化设备等。云计算虚拟化技术^[2]的出现使得运行在虚拟机上的服务器能够很方便地迁移。将云计算网络中的数据中心抽象建模为图, 由于异构性、隐私保护、数据不完整、数据不精确等原因, 该图数据存在不确定性。对于不确定性数据处理, 目前已有很多成果, 多集中在实体数据的不确定性, 对现实中一些实际问题覆盖还不够。云计算网络上数据中心之间的距离不是一个具体的数, 而是一个不确定量, 这时就无法用传统的基于两点间有确定距离值

收稿日期: 2015-06-23; 改回日期: 2015-12-08; 网络出版: 2016-01-22

*通信作者: 郭长友 guochangyouustb@139.com

基金项目: 国家自然科学基金(61163025), 北京市重点实验室 2012 年度阶梯计划项目(Z121101002812005)

Foundation Items: The National Natural Science Foundation of China (61163025), The Project of Beijing Key Laboratory of Knowledge Engineering for Materials Science (Z121101002812005)

的算法，求得最有效近邻。此时，不确定网络中的近邻查询问题的研究就非常有必要了。对于实体间关系的非确定性处理，文献[3,4]运用随机和模糊理论解决近邻查询处理问题。

实际上，在上述文献中还是用概率、模糊刻画现实中的实体间关系，不能科学地反映被判断对象的真实情况。而实体间关系有时还表现为主观不确定性，这种主观不确定性既不是随机的也不是模糊的，如云计算网络上数据中心内部虚拟机之间的动态迁移性，使得数据在迁移前考虑源和目的节点间的距离、成本、代价等数据时具有很大的主观不确定性。再者，模糊图中两点间的最短距离的可信度用求和来计算也存在问题。因为可信度的取值范围应该是在[0,1]区间，故对两点间最短距离的可信度用求和运算存在溢出等不合理现象，建立在模糊理论基础上的这些方法都不能从根本上解决上述问题。据我们所知，基于不确定理论的不确定网络条件可信近邻查询尚未被研究过。现实中，很多问题无法获得历史数据，从而无法用概率论求解事件发生的频率，此时必须依据专家经验对事件可能发生的信度进行评估，此方法使得信度的方差远远大于频率。为了处理主观不确定性，文献[5-7]于2007年创立了数学的一个新的分支——不确定理论。之后不确定理论被广泛应用于科学和工程中，解决了许多多的问题^[8-16]。

本文是运用不确定理论对复杂云计算环境下数据中心不确定网络的近邻查询问题进行分析、评价和测试，提出了不确定网络在信度、距离等条件约束下近邻查询新的解决方法，理论分析与实验结果表明，不确定网络条件可信近邻查询算法能够从不确定角度解决不确定网络环境下的近邻查询问题，且符合实际情况。

2 相关概念

2.1 不确定理论

不确定理论是清华大学刘宝碇教授在2007年^[5]提出并于2010年^[6]进行了修订，为处理不确定因素提供了一种新的研究方法。

现在，我们介绍一些本文所用的不确定理论的概念和结果^[6]。

定义 1 假设 Γ 为非空集合， L 是 Γ 上的 σ -代数。任意一个元素 $A \in L$ 被称为一个事件。如果集函数 $M\{A\}$ 满足公理 1~公理 3，则称为不确定测度。

公理 1(规范性) $M\{\Gamma\} = 1$ 。

公理 2(对偶性) 对于任意事件 A 都有 $M\{A\} + M\{A^c\} = 1$ ，记 A^c 为 A 的补集。

公理 3(次可列可加性) 对任意可数的事件序列 $\{A_i\}$ ，都有 $M\left\{\bigcup_{i=1}^{\infty} A_i\right\} \leq \sum_{i=1}^{\infty} M\{A_i\}$ 。

三元组 (Γ, L, M) 被称为一个不确定空间。不确定变量 ξ 是指从不确定空间 (Γ, L, M) 到实数集上的一个测度函数。2009年刘宝碇教授^[17]定义了乘积不确定测度，得到下面的公理 4。

公理 4(乘积公理) 设 (Γ_k, L_k, M_k) 为不确定空间， $k = 1, 2, \dots$ 。则乘积不确定测度 M 是乘积 σ -代数 $L = L_1 \times L_2 \times \dots \times L_n \times \dots$ 上的不确定测度，满足 $M\left\{\prod_{k=1}^{\infty} A_k\right\} = \prod_{k=1}^{\infty} M\{A_k\}$ ，这里 A_k 是 L_k 中的任意闭事件， $k = 1, 2, \dots$ 。

如果对任意的实 Borel 集 B_1, B_2, \dots, B_m ，满足 $M\left\{\bigcap_{i=1}^m (\xi_i \in B_i)\right\} = \prod_{i=1}^m M\{\xi_i \in B_i\}$ ，则不确定变量 $\xi_1, \xi_2, \dots, \xi_m$ 被称为是独立的。如果一个函数是从 $\{0,1\}^n$ 到 $\{0,1\}$ 的一个映射。则被称为是布尔函数。如果一个不确定变量的值是 0 或 1，则被称为布尔不确定变量。

定理 1 假设 $\xi_1, \xi_2, \dots, \xi_n$ 是独立的布尔不确定变量，即

$$\xi_i = \begin{cases} 1, & M\{\xi_i = 1\} = a_i \\ 0, & M\{\xi_i = 0\} = 1 - a_i \end{cases}$$

其中 $i = 1, 2, \dots, n$ 。如果 f 是一个布尔函数，则 $\xi = f(\xi_1, \xi_2, \dots, \xi_n)$ 是布尔不确定变量使得

$$M\{\xi = 1\} = \begin{cases} \sup_{f(x_1, x_2, \dots, x_n)=1} \min_{1 \leq i \leq n} v_i(x_i), \\ \sup_{f(x_1, x_2, \dots, x_n)=1} \min_{1 \leq i \leq n} v_i(x_i) < 0.5 \\ 1 - \sup_{f(x_1, x_2, \dots, x_n)=0} \min_{1 \leq i \leq n} v_i(x_i), \\ \sup_{f(x_1, x_2, \dots, x_n)=1} \min_{1 \leq i \leq n} v_i(x_i) \geq 0.5 \end{cases}$$

这里 x_i 取值为 0 或 1，定义 v_i 为

$$v_i(x_i) = \begin{cases} a_i, & x_i = 1 \\ 1 - a_i, & x_i = 0 \end{cases}, \quad i = 1, 2, \dots, n$$

2.2 不确定图的相关概念

集合元素之间的关系可以用两个相关元素构成的有序对来表示。这种关系可以用经典图论中的图来表示，即用结点(称之为图的顶点)表示实体，结点间的连线(称之为图的边)表示实体间的关系，若给图中的顶点或边赋予一定的信息(称之为权)，则将此图称之为网络，例如以云计算数据中心为结点构建的云计算网络，两个云计算数据中心之间的距离就可以用他们之间的边权来表示。但是现实生活中，两个实体之间是否有关系不能完全确定，有时即使关系能确定，但关系信息(即权值)却不能确定，

那么如何来刻画这种关系呢？ 本文先给出相关概念。

定义 2(不确定网络) 不确定网络是指网络实体确定而实体间关系的权值不确定， 表现为实体间关系权值的不确定性。

不确定网络中关系权值的不确定性可以是某种关系权值是否存在， 如云计算网络中以参与计算的数据中心之间的距离为权， 也可以是某种关系权值的取值情况以不同的不确定测度存在， 如云计算网络中以参与计算的数据中心之间的流量速度为权， 本文所指的不确定网络均为前者， 后者将另行撰文。

定义 3^[12](不确定图) 设 $G = (V, E, M)$ 表示不确定图， 其中， V 表示图的结点集； E 表示图的边集； M 表示边存在的不确定测度集。 $M(e)$ 表示边 e 存在的不确定测度。 其中 $0 \leq M(e) \leq 1$ ， $M(e) = 1$ 表示边 e 一定存在， $M(e) = 0$ 表示边 e 一定不存在， $0 < M(e) < 1$ 称为不确定边。

我们将网络实体确定， 实体间关系的不确定性用不确定图来描述。 结点表示网络实体， 边表示实体间关系， 边的不确定测度表示实体间关系的可信度。 经典图论中的图可以用它的邻接矩阵表示， 不确定图也可以用它的邻接矩阵表示。 从不确定图的定义中， 我们可以得出这样的结论： 不确定图 G 的边集中每个元素都是一个不确定布尔变量， 其权值是此元素取值为 1 时的不确定测度。 即 $E(G) = \{\xi_{12}, \xi_{13}, \dots, \xi_{1n}, \xi_{23}, \dots, \xi_{2n}, \dots, \xi_{(n-1)n}\}$ ， 其中 $M\{\xi_{ij}=1\} = a_{ij}$ ， $1 \leq i < j \leq n$ 。 为简单起见， 移除满足 $M\{\xi_{ij}=1\} = 0$ 的边 ξ_{ij} ， 而将边集表示为 $E(G) = \{\xi_1, \xi_2, \dots, \xi_m\}$ 。

不确定图可分为有向图与无向图， 加权图与无权图。 这里讨论的是无向加权不确定图(即不确定网络)。 图 1 为不确定图及其邻接矩阵示例。

在图 1 中， 设不确定图 $G = (V, E, M)$ ， 其中， $V = \{v_1, v_2, v_3, v_4\}$ ， $E = \{v_1v_2, v_1v_4, v_2v_3, v_3v_4\}$ ， $M = \{M(v_1v_2), M(v_1v_4), M(v_2v_3), M(v_3v_4)\}$ ， 它有 3 条不确定边 v_1v_4, v_2v_3, v_3v_4 ， 其不确定测度分别为 0.8, 0.3 和 0.4。

定义 4(样本图) 已知不确定图 $G = (V, E, M)$ ， 样本图 G_k 是不确定图 G 的一个实例， 一个有 m

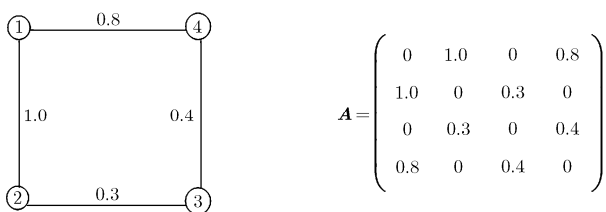


图 1 不确定图及其邻接矩阵示例

条不确定边的不确定图有 2^m 个样本图。

图 2 是图 1 所示不确定图 G 的样本图示例， 因图 1 中的不确定图有 3 条不确定边， 故其有 8 个样本图， 其中， 图 2(a) 为样本图 $G_1, E_{G_1} = \{v_1v_2\}$ ， 图 2(h) 为样本图 $G_8, E_{G_8} = \{v_1v_2, v_1v_4, v_2v_3, v_3v_4\}$ 。

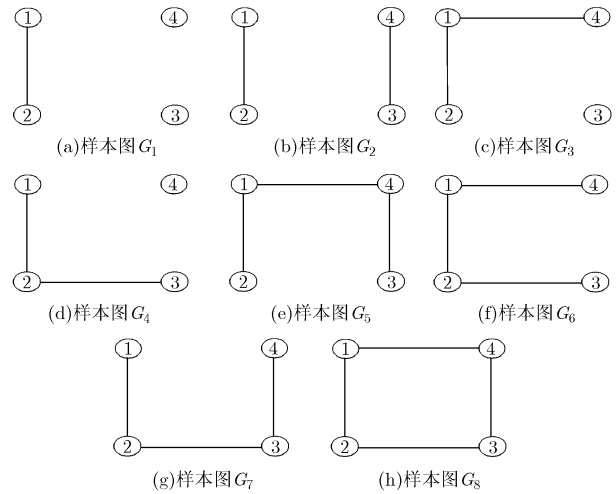


图 2 图 1 的样本图示例

为了说明不确定图是某个样本图， 下面给出样本图指数的定义。

定义 5 假设图 G 是一个不确定图， 定义图 G 的样本图函数为

$$S_k(G) = \begin{cases} 1, & \text{不确定图 } G \text{ 是样本图 } G_k \\ 0, & \text{其它} \end{cases}, k = 1, 2, \dots, 2^m$$

显然， $S_k(G)$ 是一个布尔函数。 一个不确定图 G 的样本图指数被定义为

$$\rho_k(G) = M\{S_k(G) = 1\}, k = 1, 2, \dots, 2^m$$

也就是说， 不确定图 G 的样本图指数是指图 G 是某个样本图 G_k 的不确定测度。

关键的问题是当给出了一个不确定图时， 如何获得它的样本图指数。 下面的定理解决了这个问题。

定理 2 假设 G 是一个 n 阶不确定图其邻接矩阵为

$$A = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{n1} & \alpha_{n2} & \dots & \alpha_{nn} \end{pmatrix}$$

如果所有的边都是独立的， 则 G 的样本图指数为

$$\rho_k(G) = \begin{cases} \sup_{S_k(G)=1} \min_{1 \leq i, j \leq n} v_{ij}(x_{ij}), \\ \sup_{S_k(G)=1} \min_{1 \leq i, j \leq n} v_{ij}(x_{ij}) < 0.5 \\ 1 - \sup_{S_k(G)=0} \min_{1 \leq i, j \leq n} v_{ij}(x_{ij}), \\ \sup_{S_k(G)=1} \min_{1 \leq i, j \leq n} v_{ij}(x_{ij}) \geq 0.5 \end{cases}, k = 1, 2, \dots, 2^m$$

其中 \mathbf{X} 是 $n \times n$ 不确定对称矩阵且 \mathbf{X} 的主对角线元素是零使得

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{pmatrix}$$

这里 $x_{ji} = x_{ij}$, 其值为0或1, v_{ij} 的取值为

$$v_{ij}(x_{ij}) = \begin{cases} a_{ij}, & x_{ij} = 1 \\ 1 - a_{ij}, & x_{ij} = 0, \end{cases} \quad i, j = 1, 2, \dots, n$$

证明 因为 x_{ij} 是独立的布尔型不确定变量 ($i = 1, 2, \dots, n; j = 1, 2, \dots, n$), 则 \mathbf{X} 是 $n \times n$ 布尔不确定对称矩阵。可知 $S_k(G)$ 是布尔函数。根据定理 1, 可得

$$M(S_k(G) = 1) = \begin{cases} \sup_{S_k(G)=1} \min_{1 \leq i, j \leq n} v_{ij}(x_{ij}), \\ \sup_{S_k(G)=1} \min_{1 \leq i, j \leq n} v_{ij}(x_{ij}) < 0.5 \\ 1 - \sup_{S_k(G)=0} \min_{1 \leq i, j \leq n} v_{ij}(x_{ij}), \\ \sup_{S_k(G)=1} \min_{1 \leq i, j \leq n} v_{ij}(x_{ij}) \geq 0.5 \end{cases}$$

其中 \mathbf{X} 是 $n \times n$ 不确定对称矩阵且 \mathbf{X} 的主对角线元素是零使得

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{pmatrix}$$

这里 $x_{ji} = x_{ij}$, 其值为0或1, 而 $x_{ii} = 0, v_{ij}$ 的取值为

$$v_{ij}(x_{ij}) = \begin{cases} a_{ij}, & x_{ij} = 1 \\ 1 - a_{ij}, & x_{ij} = 0, \end{cases} \quad i, j = 1, 2, \dots, n$$

根据定义5, 不确定图的样本图指数是指这个图是此样本图的不确定测度。因此定理2得证。

设不确定赋权图 $G = (V, E, W, M)$, 其中, V 表示图的结点集; E 表示图的边集; W 表示边的权重集; M 表示边存在的不确定测度集。 G_k 是 G 的赋权样本图, $v_i, v_j \in V$, 图中两点间连线的权重 $w_{ij} \in W$ 表示两点 v_i, v_j 间的直接距离。图 3 为建模为图 1 的不确定网络示例, 其中括号内的数字表示边的权值, 若两结点间无边相连, 则定义其权为无穷大。图 4 为图 3 的赋权样本图。

定义 6 赋权样本图中两点间路径的权是指这条路径上所有边的权和, 赋权样本图中两点间最短路径是指这两点间所有路径中权最小的路径, 最小路径的权称为这两点之间的距离, 记为 $d_{G_k}(v_i, v_j) = d$ 表示 v_i, v_j 在 G_k 中的距离为 d , $d_{G_k}(v_i, v_j)$ 为 d 的可信度为 $M(d_{G_k}(v_i, v_j) = d) = \rho_k(G)$ 。

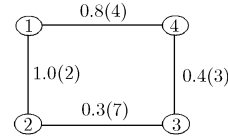


图 3 不确定网络示例

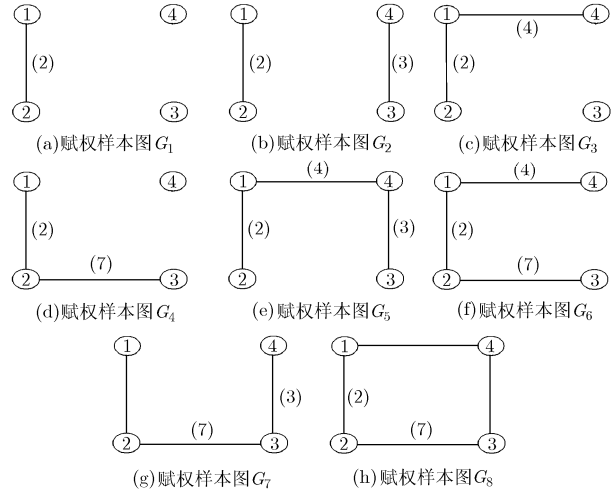


图 4 图 3 的赋权样本图示例

定义 7(不确定网络中两点间距离) 不确定网络 G 中 v_i, v_j 间距离记为 $d(v_i, v_j)$, $d(v_i, v_j)$ 为 d 的可信度为: $M(d(v_i, v_j) = d) = \max_{1 \leq k \leq 2^m} M(d_{G_k}(v_i, v_j) = d)$, 其中 m 为不确定图 G 中不确定边的条数。

定义 8(可信距离) 不确定网络中 v_i, v_j 间可信距离 $d_M(v_i, v_j) = \arg(\max M(d(v_i, v_j) = d))$ 。

2.3 可信距离的计算方法

可信距离查询算法如表 1 所示。

表 1 可信距离查询算法

<p>算法 1 不确定网络中任意两点间的可信距离查询算法</p> <p>输入: 不确定网络 $G = (V, E, W, M)$, 任意查询结点 $v_p, v_q \in V$。</p> <p>输出: 不确定网络中查询结点 v_p, v_q 间的可信距离 $d_M(v_p, v_q)$。</p> <ol style="list-style-type: none"> (1)先将不确定网络建模为不确定图, 求得不确定图的邻接矩阵; (2)设不确定图有 m 条边, 求得 2^m 个样本图的邻接矩阵; (3)由不确定图的邻接矩阵和样本图的邻接矩阵及样本图指数的定义可计算得出所有样本图指数; (4)找出不确定网络的所有赋权样本图; (5)在每个样本图 G_k 中求出查询结点 v_p, v_q 间的可信距离 $d_{G_k}(v_p, v_q)$ 及其可信度 $M(d_{G_k}(v_i, v_j) = d) = \rho_k(G)$; (6)计算出不确定网络中任意结点 v_i, v_j 间可信距离 $d_M(v_i, v_j) = \arg(\max M(d(v_i, v_j) = d))$; (7)求得结点 v_p, v_q 间的可信距离 $d_M(v_p, v_q)$。

不确定网络图中任意两点间可信距离的精确计算时间复杂度较高，达到指数级。在现实生活中，由于各种因素的限制，往往并不追求最优解，只要求得一个满意解即可，因此采用以下两种方法进行近似计算。

定义 9(基础网络) 已知不确定网络 $G = (V, E, W, M)$ 的基础网络 G' 是指不确定网络 G 中将不确定测度集 M 中 $0 < M(e) \leq 1$ 的元素映射为 1 的一个确定网络。图 5 为图 3 的基础网络示意图。

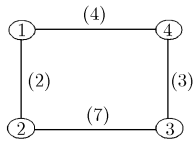


图 5 图 3 的基础网络

定义 10(可达路径长度) 不确定网络中 v_i, v_j 间可达路径长度是指从 v_i 到 v_j 的简单路径所经过的所有边的权和，记为 $l(v_i, v_j)$ ；其不确定测度为这条路径中所有边的不确定测度的最小值，记为 $M_l(v_i, v_j)$ ，即

$$l(v_i, v_j) = \sum_{e \in l} W(e), M_l(v_i, v_j) = \min_{e \in l} M(e)$$

定义 11(可达路径期望长度) v_i, v_j 间可达路径期望长度为 $L_{ME}(v_i, v_j) = \sum l \times M(l(v_i, v_j) = l)$ 。

2.4 可达路径长度计算

可达路径长度计算算法如表 2 所示。

表 2 可达路径长度计算算法

<p>算法 2 不确定网络中任意两点间的可达路径长度算法</p> <p>输入：不确定网络 $G=(V, E, W, M)$，任意查询结点 $v_p, v_q \in V$。</p> <p>输出：不确定网络中查询结点 v_p, v_q 间的可达路径长度 $l(v_p, v_q)$。</p> <p>(1)找出不确定网络 $G = (V, E, W, M)$ 的基础网络 G'；</p> <p>(2)在基础网络中找出任意两点 v_i, v_j 间的所有简单路径，并计算其长度；</p> <p>(3)根据定义 6，计算出基础网络中找出的任意两点间的所有简单路径的信度；</p> <p>(4)求得查询结点 v_p, v_q 间的可达路径长度 $l(v_p, v_q)$。</p>

2.5 可达路径期望长度计算

可达路径期望长度计算算法如表 3 所示。

3 条件可信 k 近邻查询

定义 12(条件可信 k 近邻查询) 在不确定网络 $G = (V, E, W, M)$ 中，给定结点 $v_q \in V$ ，条件可信 k 近邻查询定义为查询出与 v_q 的可达路径长度(或可达路径期望长度)小于要求长度且其可信度大于要求信度的 V 中 k 个对象组成的结点集。

表 3 可达路径期望长度计算算法

<p>算法 3 不确定网络中任意两点间的可达路径期望长度算法</p> <p>输入：不确定网络 $G = (V, E, W, M)$，任意查询结点 $v_p, v_q \in V$。</p> <p>输出：不确定网络中查询结点 v_p, v_q 间的可达路径期望长度 $l_{ME}(v_p, v_q)$。</p> <p>(1)找出不确定网络 $G = (V, E, W, M)$ 的基础网络 G'；</p> <p>(2)在基础网络中找出任意两点 v_i, v_j 间的所有简单路径，并计算其长度；</p> <p>(3)根据定义 6，计算出基础网络中找出的任意两点间的所有简单路径的信度；</p> <p>(4)根据定义 11，计算出不确定网络中任意两点 v_i, v_j 间可达路径期望长度 $l_{ME}(v_i, v_j)$；</p> <p>(5)求得查询结点 v_p, v_q 间的可达路径期望长度 $l_{ME}(v_p, v_q)$。</p>
--

条件可信 k 近邻查询算法如表 4 所示。

表 4 条件可信 k 近邻查询算法

<p>算法 4 条件可信 k 近邻查询算法</p> <p>输入：不确定网络 $G=(V, E, W, M)$，任意查询结点 $v_p, v_q \in V$，约束条件可达路径长度 $l(v_p, v_j)$ (或可达路径期望长度 $l_{ME}(v_p, v_j)$)，约束条件可信度 $M(v_p, v_j)$。</p> <p>输出：满足与任意查询结点 v_p 的条件约束，即与 v_p 的可达路径长度 $l(v_p, v_j) \leq L$ (或可达路径期望长度 $l_{ME}(v_p, v_j) \leq L$) 小于要求长度且其可信度大于要求信度 $M_l(v_p, v_j) \geq M'$ 的 V 中 k 个结点集合。</p> <p>条件可信近邻查询计算方法的主要步骤如下：</p> <p>(1)将不确定网络 $G = (V, E, W, M)$ 中，信度小于已知要求信度 M' 的不确定边删除(或将其信度降为 0)，建立新的不确定网络 $G' = (V', E', W', M')$；</p> <p>(2)在 G' 的基础网络中找出任意两点 v_i, v_j 间的所有简单路径，并计算其长度；</p> <p>(3)在 G' 的基础网络中找出与 v_p 点最近的，距离小于 L 的 k 个点。</p>
--

下面给出一个云计算数据中心资源管理中虚拟机迁移的例子来说明不确定网络中条件可信近邻查询的应用。在虚拟化云计算资源的管理过程中将资源负载结点抽象为不确定网络的结点，将资源负载异常结点和待迁移目标结点间的可迁移性抽象为不确定网络中不确定边的信度，将迁移过程所消耗的时间抽象为不确定网络中不确定边的权重。如图 6 所示不确定网络近邻查询示例图中，共有 16 个节点，30 条边，其中数字 i 表示节点 v_i ，边 v_1, v_2 的权为 1，可信度为 0.1；边 v_2, v_3 的权为 2，可信度为 0.3；边 v_3, v_4 的权为 3，可信度为 0.6；其余边的权及可信度如图 6 所示。

要求查询出与 v_1 的可达路径长度(或可达路径

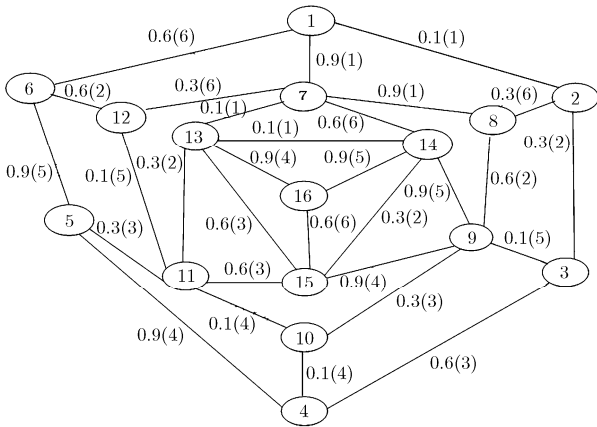


图 6 不确定网络近邻查询示例

期望长度)小于5且其可信度大于0.5的最近的3个近邻接点。

由条件可信近邻查询计算方法,按如下步骤计算:

步骤 1 将信度小于0.5的不确定边删除,建立新的不确定网络,如图7所示;

步骤 2 找出信度大于0.5的新不确定网络的基础网络,如图8所示;

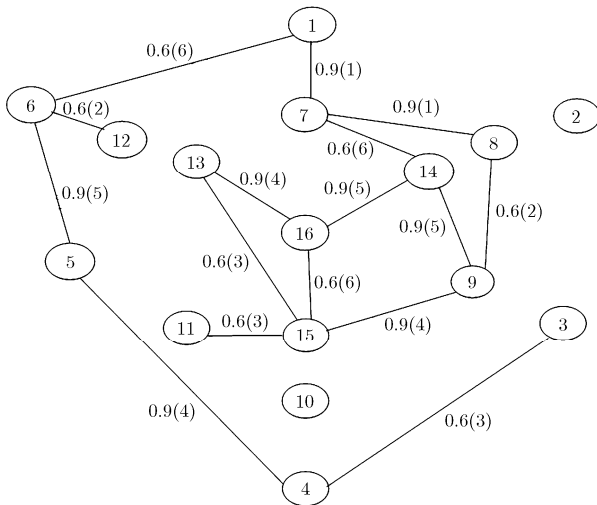


图 7 删除不满足信度条件不确定边后的不确定网络

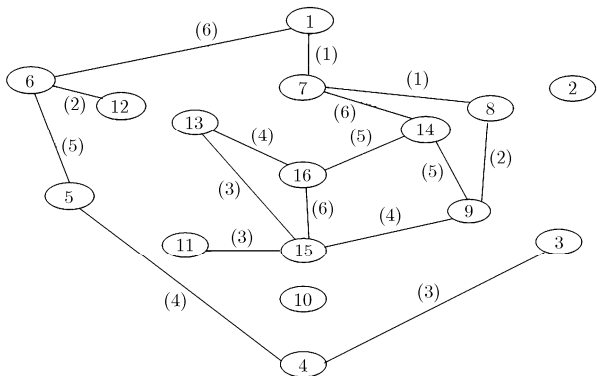


图 8 图7的基础网络

步骤 3 在信度大于0.5的新不确定网络的基础网络中找出与结点 v_1 可达路径长度(或可达路径期望长度)小于5的最近3个点。

由定义6~定义10可计算得出在信度大于0.5的约束下, v_1 到 v_2 有0条可达路径, v_1 到 v_3 有1条可达路径, 可达路径长度为18, v_1 到 v_4 有1条可达路径, 可达路径长度为15, 同理可求得 v_1 到其余结点的可达路径数及可达路径长度。与结点 v_1 可达路径长度小于5的, 最近的3个点是 v_7, v_8, v_9 , 对应的可达路径长度分别是1, 2, 4。由定义6~定义11可计算得出 v_1 到 v_2 可达路径期望长度为0, v_1 到 v_3 可达路径期望长度为10.8, v_1 到 v_4 可达路径期望长度为9, 同理可求得 v_1 到其余结点的可达路径期望长度。与结点 v_1 可达路径期望长度小于5, 最近的3个点是 v_6, v_7, v_{12} , 对应的可达路径期望长度分别是3.6, 0.9, 4.8。对于上述条件可信 k 近邻查询求解过程中, 可能会出现满足可信度、可达路径长度(或可达路径期望长度)约束的节点多于 k 个对象的情况, 这时可以对求得的节点集合按信度优先或可达路径长度(或可达路径期望长度)优先选取 k 个节点。

4 实验结果分析

下面通过实验来验证本文所提出的不确定网络条件可信近邻查询算法的执行效率及其稳定性。由于目前尚没有处理不确定网络中基于信度、可达路径期望长度等条件约束下的可信近邻查询算法, 因此本文主要考察提出的查询方法在不同规模的数据集上以及不同约束条件下的效率。

实验数据在 Windows 7 操作系统上, 1.6 GHz 处理器和 4 GB 内存条件下, 用 Visual C++6.0 编程环境采用人工模拟不确定网络实现。

实验使用定义 2 中定义的不确定网络。随机生成 T1, T2, ..., T10 10 个不确定网络。每个不确定网络的节点和不确定边的数量如表 5 所示。每个不确定网络的不确定边随机在两个点间生成, 每个不确定边的信度为随机生成的 [0,1] 间的实数, 每个不确定边的权重为随机生成的 [0,1000] 间的实数。

实验分为两部分, 第 1 部分测试可信度对基于可信距离的近邻查询算法的影响, 第 2 部分测试可达路径期望长度对基于可信距离的近邻查询算法的影响。

实验 1 测试可信度对基于可信距离的近邻查询算法的影响, 可信度的查询结果如图 9 所示。从图 9 可知, 可信度对基于可信距离的近邻查询算法有相似分布, 随着不确定网络中节点和不确定边的数量变化而相应变化。

表 5 实验使用人工模拟的不确定网络数据

不确定图名称	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
节点数	140	130	120	110	100	90	80	70	60	50
不确定边数	9000	7000	5000	5000	4500	3500	3000	2000	1700	700

实验 2 测试可达路径期望长度对基于可信距离的近邻查询算法的影响。可达路径期望长度的查询结果如图 10 所示。从图 10 可知，可达路径期望长度对基于可信距离的近邻查询算法情况相似，有相似分布，随着不确定网络中节点和不确定边的数量变化而相应变化。

5 结束语

本文研究了不确定网络条件可信近邻查询问题，首次运用不确定理论形式化定义了不确定网络条件可信近邻查询问题，提出了基于不确定理论的不确定网络、不确定图及其邻接矩阵、样本图及其指数、不确定网络中两点间距离、可信距离的定义，

设计了不确定网络中任意节点间可信距离的求解算法。鉴于不确定网络中，节点间可信距离的精确计算代价较高，本文又提出了不确定网络的基础网络、可达路径长度、可达路径期望长度的定义，在此基础上给出了信度，可达路径长度(或可达路径期望长度)约束条件下的高效可信近邻查询算法。本文进行了大量实验来考察该算法的性能。

理论分析与实验结果表明算法可行且性能稳定，能有效处理不确定网络条件可信近邻查询。在不确定网络中同时考虑不确定性与随机性，使非确定网络环境下的不确定随机近邻查询更有效是下一步的研究方向。

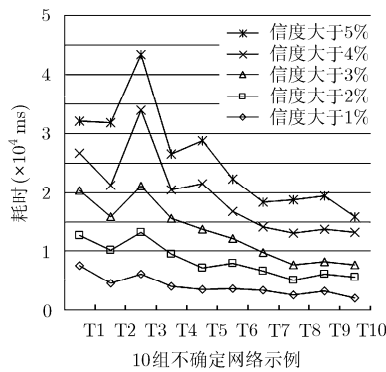


图 9 可信距离分布

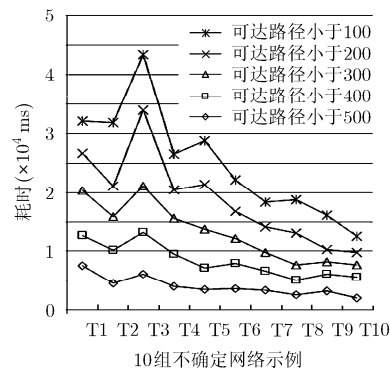


图 10 可信距离与样本数关系

参考文献

- [1] 罗亮, 吴文峻, 张飞. 面向云计算数据中心的能耗建模方法[J]. 软件学报, 2014, 25(7): 1371-1387. doi: 10.13328/j.cnki.jos.004604.
- [2] LUO L, WU W J, and ZHANG F. Energy modeling based on clouddata center[J]. *Journal of Software*, 2014, 25(7): 1371-1387. doi: 10.13328/j.cnki.jos.004604.
- [3] 殷波, 王颖, 邱雪松, 等. 一种面向云服务提供商的资源分配机制[J]. 电子与信息学报, 2014, 36(1): 15-21. doi: 10.3724/SP.J.1146.2013.00427.
- [4] YIN Bo, WANG Ying, QIU Xuesong, et al. A resource provisioning mechanism for service providers in cloud[J]. *Journal of Electronics & Information Technology*, 2014, 36(1): 15-21. doi: 10.3724/SP.J.1146.2013.00427.
- [5] 张海杰, 姜守旭, 邹兆年. 不确定图上的高效top-k近邻查询处理算法[J]. 计算机学报, 2011, 34(10): 1885-1896. doi: 10.3724/SP.J.1016.2011.01885.
- [6] ZHANG Haijie, JIANG Shouxu, and ZOU Zhaonian. An efficient algorithm for top-k proximity query on uncertain graph[J]. *Chinese Journal of Computers*, 2011, 34(10): 1885-1896. doi: 10.3724/SP.J.1016.2011.01885.
- [7] 高峻, 郝忠孝. 受限模糊网络可信近邻查询[J]. 计算机工程, 2015, 41(1): 54-60. doi: 10.3969/j.issn.1000-3428.2015.01.010.
- [8] GAO Jun and HAO Zhongxiao. Credible nearest neighbor query in constraint fuzzy network[J]. *Computer Engineering*, 2015, 41(1): 54-60. doi: 10.3969/j.issn.1000-3428.2015.01.010.
- [9] LIU B. Uncertainty Theory[M]. 2nd ed., Berlin: Springer-Verlag, 2007, Chapter 1-Chapter 2.
- [10] LIU B. Uncertainty Theory: A Branch of Mathematics for Modeling Human Uncertainty[M]. Berlin: Springer-Verlag, 2010, Chapter 1-Chapter 2.

- [7] LIU B. Uncertainty distribution and independence of uncertain processes[J]. *Fuzzy Optimization and Decision Making*, 2014, 13(3): 259–271. doi: 10.1007/s10700-014-9181-5.
- [8] ZHOU J, CHEN L, and WANG K. Path optimality conditions for minimum spanning tree problem with uncertain edge weights[J]. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 2015, 23(1): 49–71. doi: 10.1142/S0218488515500038.
- [9] GAO X L. Uncertain relations on a finite set and their properties[J]. *Pure and Applied Mathematics Journal*, 2014, 3(1): 13–19. doi: 10.11648/j.pamj.s.20140301.13.
- [10] GAO X L. Tree index of uncertain graphs[J]. *Soft Computing*, 2015. doi: 10.1007/s00500-015-1597-5.
- [11] GAO X L and GAO Y. Connectedness index of uncertainty graphs[J]. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 2013, 21(1): 127–137. doi: 10.1142/S0218488513500074.
- [12] GAO X L. Regularity index of uncertain graph[J]. *Journal of Intelligent & Fuzzy Systems*, 2014, 27(4): 1671–1678. doi: 10.3233/IFS-141133.
- [13] DING S B. Uncertain minimum cost flow problem[J]. *Soft Computing*, 2014, 18(11): 2201–2207. doi: 10.1007/s00500-013-1194-4.
- [14] Gao X, Gao Y, and Ralescu D. On Liu's inference rule for uncertain systems[J]. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 2010, 18(1): 1–11. doi: 10.1142/S0218488510006349.
- [15] Gao Y, Yang L X, et al. On distribution function of the diameter in uncertain graph[J]. *Information Sciences*, 2015, 296(1): 61–74. doi: 10.1016/j.ins.2014.10.048.
- [16] GAO Y. Shortest path problem with uncertain arc lengths[J]. *Computers and Mathematics with Applications*, 2011, 62(6): 2591–2600. doi: 10.1016/j.camwa.2011.07.058.
- [17] LIU B. Some research problems in uncertainty theory[J]. *Journal of Uncertain Systems*, 2009, 3(1): 3–10.

郭长友：男，1976年生，副教授，研究方向为网络安全。

郑雪峰：男，1951年生，教授，研究方向为计算机系统安全、网络安全。

高秀莲：女，1969年生，副教授，研究方向为图论与组合优化。