

A Data Aggregation Algorithm Based on Grid and Mobile Agent in WSN*

LI Zhi-yu^{1,2*}, SHI Haoshan²

1. Department of Mechatronics, Northwestern Polytechnical University, Xi'an 710072, China;
2. Department of Electronic Engineering, Northwestern Polytechnical University, Xi'an 710072, China

Abstract: Energy efficiency is an important concern in the routing algorithm design for Wireless Sensor Networks (WSN). Data aggregation can efficiently decrease energy consumption by aggregating the redundant information. This paper presents a data aggregation algorithm that integrates grid generation and mobile agent (MA) in WSN, which is based on MA retransmitting interest agent packets sent by sink node and data agent packets sent by target nodes. In this algorithm, MA routing can be treated as an optimization problem. WSN is evenly divided into many two-dimensional grids, forming the initial population of adaptive genetic algorithm (AGA). AGA is applied to find out the optimal nodes' sequence of MA routing. Simulation results show that compared with Local Closest First algorithm (LCF), with the increasing of the network scale, the proposed algorithm can produce less energy consumption and network delay.

Key words: wireless sensor networks; data aggregation; grid generation; mobile agent; adaptive genetic algorithm

EEACC:7230 G

一种基于网格和移动代理的 无线传感器网络数据融合算法*

李志宇^{1,2*}, 史浩山²

1. 西北工业大学机电学院, 西安 710072;
2. 西北工业大学电子信息学院, 西安 710072

摘要: 能源有效性是无线传感器网络(WSN)路由算法设计要考虑的首要问题,数据融合可以通过合并冗余数据而有效地节约能耗。提出一种将网格和移动代理相结合的 WSN 数据融合算法,基于移动代理对 Sink 节点发出兴趣代理报文和目标节点发出数据代理报文进行转发。将移动代理路由归结为一个优化问题,通过把 WSN 均匀分割为多个大小适当的二维网格,形成自适应遗传算法(AGA)的初始群体,采用 AGA 求出移动代理的最优路由由节点序列。仿真结果表明,随着网络规模增大,和局部最近邻优先算法(LCF)相比,该算法有更小的网络能耗和延时。

关键词: 无线传感器网络;数据融合;网格划分;移动代理;自适应遗传算法

中图分类号: TP212

文献标识码: A 文章编号: 1004-1699(2008)07-1232-05

无线传感器网络(Wireless Sensor Networks, WSN)因其高可靠、易部署和易扩展等特点逐渐受到人们的重视,并迅速发展起来。与传统的网络相比,WSN 中的每个独立节点的能量、监测范围及可靠性是有限的,为了增强整个网络所能获得信息的鲁棒性和准确性,在放置节点时必须使节点的监测范围互相交叠,这样,节点所采集到的数据就存在一定的冗余性,网络负载会随之增加。因为处理数据

所消耗的能量与传送数据所消耗的能量相比要小得多,所以应对冗余数据进行网内处理,通过数据融合平衡网络内部各节点的能量消耗,可以延长网络的生存时间。数据融合技术是 WSN 研究中的重要部分^[1-2]。

目前,WSN 的数据融合算法主要基于传统的 Client/Server(C/S)模型。在这种模型中,传感节点被看成是 Client, Sink 节点被看作是 Server, Sink 节

基金项目:国家自然科学基金资助(60472074);教育部博士点基金项目资助(20050699037)

收稿日期:2007-12-11 修改日期:2008-03-18

点收集依据路由协议从各个传感节点传送来的数据,集中进行融合处理。对于 WSN 来说,C/S 模型存在诸多问题:一是网络延迟和能量消耗大,各传感节点可以同时向处理节点传送数据,而处理节点只能顺序接收,当传感节点数量增多,传感数据总量增大时,网络延迟和能量消耗随之加大;二是可伸缩性(scalability)差,当 WSN 补充新的传感节点时,网络结构常常需要调整以保持负载平衡;三是节点能量消耗不均衡,处理节点因为要保持与各个传感节点的连接并处理其数据,相对传感节点来说要消耗更多的能量,这就需要预置超强能量的节点作为处理节点或采用某种算法来轮换处理节点,增加了额外的网络开销^[3]。随着 WSN 中采用基于移动代理(Mobile Agent, MA)的计算模型的提出,在一定程度上解决了传统 C/S 模型所带来的问题^[4]。

文献[4]提出了局部最近邻优先(Local Closest First, LCF)启发式 MA 数据融合路由算法,MA 每次选取离当前节点最近的邻居节点作为下一跳节点,在其访问的每个传感节点中进行数据融合,但它只适于简单环境下的小规模传感器网络,当网络规模不断扩大以及传感器节点的分布更复杂时,启发式算法只能求得局部最优 MA 路由。文献[5]利用遗传算法(Genetic Algorithm, GA)求解最小化网络能量的全局最优 MA 路由,但其融合能量是时间的函数而非数据量的函数,所以不能说明因数据融合而减少了数据传输量,进而减少了传输能量,达到节能目的。

本文提出了一种基于网格划分和移动代理的无线传感器网络数据融合算法(Data Aggregation Algorithm based on Grid Generation and Mobile Agent in Wireless Sensor Networks, DAGGMA),将移动代理路由归结为一个优化问题,通过把 WSN 均匀分割为多个大小适当的二维网格,形成自适应遗传算法(Adaptive Genetic Algorithm, AGA)的初始群体,采用 AGA 求出移动代理的最优路由由节点序列。仿真结果表明,随着网络规模增大,和 LCF 算法相比,该算法有更小的网络能耗和延时。

1 WSN 中的 MA 路由问题

无线传感器网络一般包括:Sink 节点、传感器节点和目标节点。其中 Sink 节点有相对较强的处理、存储和通信能力及进行全局优化的先决知识,它根据当前的网络状态,决策 MA 的最优路由。MA 路由确定 MA 在网络中迁移的待访问节点集及先后顺序,根据规划时机,MA 从 Sink 节点出发,按照事先确定的路由顺次访问传感器节点,收集并融合

感兴趣的数据,最后 MA 将携带数据融合结果回到 Sink 节点,本文只考虑单 Sink 节点的情况。在区域内访问到的节点越多,收集到的数据经过融合后所得结果准确性就越高,但是为了降低能源消费总量和路径损失的成本,必须为 MA 选择一条适宜的访问路径。MA 路由要求以最小系统代价满足应用所需性能,比如以最低能耗和最小延迟获得期望精度的数据。

定义 1 用带权重的连通图 $G = G(V, E,)$ 表示给定的 WSN,其中, V 为节点集合; E 表示链路的集合; 表示链路权重,通常用来评估两个节点间通信代价,如能量消耗、信号路径损耗或延时等。

定义 2 假定 MA 路由过程中大小 S_{MA} (S_{MA} 表示 MA 传送单位数据量的代价)不变,只提高数据空间存储的信息精度,能量消耗为链路的权重,节点 $v(i)$ 到 $v(j)$ 的代理迁移代价为:

$$(i, j) = \begin{cases} d_{ij} \cdot S_{MA}, & d_{ij} \leq R_{\max} \\ , & d_{ij} > R_{\max} \end{cases} \quad (1)$$

其中, d_{ij} 为节点 $v(i)$ 和 $v(j)$ 间的传输距离; 传输路径信号衰减因子, $2 \leq \leq 4$; R_{\max} 为节点最大传输距离。由式(1)可以看出,如果 d_{ij} 小于 R_{\max} 表示节点可达,否则不可达。在本文中,为了简化计算, $S_{MA} = 1$, $= 2$,则节点 $v(i)$ 到 $v(j)$ 的代理迁移代价可以简化为:

$$(i, j) = \begin{cases} d_{ij}^2, & d_{ij} \leq R_{\max} \\ , & d_{ij} > R_{\max} \end{cases} \quad (2)$$

即节点 $v(i)$ 到 $v(j)$ 的代理迁移的能耗只与节点间的传输距离的平方成正比。

定义 3 节点对目标的感知过程是其接收目标发射信号的过程,定义节点 $v(i)$ 对目标感知的信息增益定义为:

$$SE(i) = \begin{cases} d_{io}, & d_{io} \leq D_{\max} \\ 0, & d_{io} > D_{\max} \end{cases} \quad (3)$$

其中, d_{io} 为节点 $v(i)$ 到目标的距离; 为目标信号传输衰减因子,在本文中,取 $= 2$; D_{\max} 为对目标最大检测距离,由式(3)可以看出, d_{io} 小于 D_{\max} 时,节点可感知目标,否则不能感知目标。

定义 4 MA 在区域内访问到的节点越多,收集到的数据经过融合后所得结果准确性就越高。但是,MA 路由要求访问路径应具有尽可能高的数据融合精度而传输代价尽可能小。出于实现和计算量考虑,定义 MA 路由问题归结为一个优化问题:

$$\min_{i, j \in V \text{ and } i \neq j} (i, j) x_{ij} \quad (4)$$

其中, x_{ij} 表示节点连通状态,“1”表示该节点可以被访问,“0”该节点不能被访问; (i, j) 由式(2)确定。

2 基于网格划分和移动代理的传感器网络数据融合算法 :DAGGMA

为了让目标节点知道 Sink 节点的位置和所需数据,算法需要进行兴趣散布。Sink 节点发出兴趣代理报文 (Interest Agent Packets, IAP) 通告其位置和感兴趣的数据类型;目标节点便据此向 Sink 节点发出数据代理报文 (Data Agent Packets, DAP) 告知其收集到的数据。DAP 和 IAP 转发通过 MA 迁移的最优路由由节点序列进行。本文采用 A GA 构造 MA 路由。

GA 是一种模拟生物进化过程的全局收敛算法,通过种群搜索策略和种群中个体间的信息互换,实现由个体组成的种群的整体学习,标准 GA 从初始种群出发,以个体适应度函数为指导,通过随机选择、交叉和变异等一系列操作,使种群逐代进化直至满足终止条件或达到最大迭代次数。由于 GA 中交叉概率和变异概率取固定值,交叉概率与变异概率不能反映进化的过程,从而容易产生早熟收敛和收敛速度慢的缺陷,而在 A GA 中,个体的交叉概率和变异概率能够随群体及个体适应度的改变自适应调整,因而能够在进化过程中保持群体多样性的同时,保证了算法的收敛性。基于 A GA 的 MA 路由算法算法设计如下:

2.1 产生初始种群

初始种群的遍历性和多样性对 GA 的性能起到关键性的影响。在标准 GA 中,初始种群内的个体通常是基于概率来随机选取,这种选择方式会使其个体缺乏遍历性和多样性。本文提出一种用网格法来产生初始的群体,增加个体的遍历性和多样性。基于文献[6],本文采用了一种基于网格划分的模型产生初始种群,其主要思想是:把 WSN 均匀分割为多个边长为 r 二维正方形网格,网格中所有节点都可以和相邻网格中所有节点通信,构造以 Sink 节点为坐标原点的网格结构,如图 1 所示,然后对网格内的节点进行编码,形成 A GA 的初始群体。

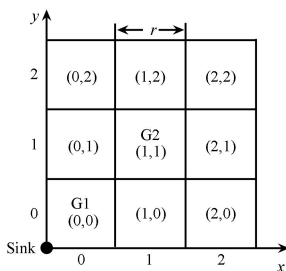


图 1 网格划分法示意图

假设网络中所有节点的通信半径 R 相同且知道自己的位置、ID 以及能量值。节点一旦被散布到

一定的网络区域内,即发送 HELLO 报文确定自己一跳范围内节点的相关信息并据此建立邻居表,HELLO 报文中包括节点位置、ID 和网格独立标志等。网格独立标志用来判定节点是否和邻居节点属于同一网格,0 为同一网格,1 为不同网格。因为 HELLO 报文非常小,网格建立期间整个网络开销也非常小。网络中的每个节点根据如下公式确定自己所在的网格 (C_x, C_y):

$$\begin{cases} C_x = \left\lfloor \frac{x_{\text{sink}}}{r} \right\rfloor \\ C_y = \left\lfloor \frac{y_{\text{sink}}}{r} \right\rfloor \end{cases} \quad (5)$$

其中, x_{sink} 和 y_{sink} 分别表示节点距离 Sink 节点的横坐标和纵坐标。网格建立后,每个节点都将知道自己所属的网格,例如 G_2 内节点所在的网格由 (1,1) 表示。为了保证相邻网格中节点之间可以相互通信,以节点的通信半径 R 作为相邻网格中任意量节点间的最远距离,则 r 和 R 的关系应满足: $r^2 + (2r)^2 \leq R^2$, 即 $r \leq R/\sqrt{5}^{[7]}$ 。实验表明,当网格较大时,单位网格内节点数增多,与其他网格内的节点交换信息次数增多会引起节点能耗增加;当网格较小时,因为网格内节点的减少,每条路径可维持的生存时间有所下降,从而造成网络生存时间减少。因此,当选取的网格过大或过小时,都会降低算法的性能,本文选取网格边长 $r = R/2.5^{[8]}$ 。

2.2 编码方式

对个体(网格)内的所有节点采用两层编码。第一层编码是 MA 路径上依次经过节点的 ID 号的排列,第二层编码为二进制编码,“1”表示该节点可以被访问,“0”该节点不能被访问。对于 IAP 转发路径,可行解应满足第一层编码序列从 Sink 节点出发,第二层二进制序列的第一位为“1”,同时采集的信号能量大于门限。对于 DAP 转发路径,可行解的第一层编码应从 IAP 转发路径最后访问的节点出发,并返回 Sink 节点,相应的第二层二进制序列码为“1”。

2.3 适应度函数

为简化讨论,最优路径的解可表示为 MA 的节点访问序列 P ,融合过程定义为累加各节点对目标的信号增益,要求累计信号增益大于预定门限值。定义适应度函数如下^[9]:

$$f(P) = \text{con} - (J(P) + g(P)) \quad (6)$$

其中, con 为较大常数以确保适应值非负; $J(P)$ 是 MA 沿路径 P 访问的代价,由式(4)决定; $g(P)$ 是惩罚项:

$$g(P) = \begin{cases} 0, & SE(i) \geq SE_T \\ i^P \\ i^P \\ (SE(i) - SE_T), & SE(i) < SE_T \end{cases} \quad (7)$$

其中, SE_T 为累计信号增益门限值; ρ 为惩罚系数, 可由实验调节。在本文中, SE_T 为目标发射信号总和的 70%, $\rho = 0.6$ 。

2.4 交叉和变异

采用赌盘方式选择用于交叉的个体。采用两点交叉作用于二层编码, 以一定的交叉概率 p_c 进行交叉操作; 防止种群陷于局部极值, 采用两点变异和反转两种变异方法作用于两层编码, 以一定的变异概率 p_m 进行变异操作。 p_c 和 p_m 对遗传算法性能的影响很大, 本文采用文献 [10] 中的自适应调整 p_c 和 p_m 方法, 其表达式为:

$$p_c = \begin{cases} \frac{p_c}{1 - \frac{f_{\max}}{f_{\min}}} & \text{if } \frac{f_{\max}}{f_{\min}} > a, \frac{f_{\min}}{f_{\max}} > b \text{ and } p_c < 1 - b \\ p_c & \text{otherwise} \end{cases} \quad (8)$$

$$p_m = \begin{cases} \frac{p_m}{1 - \frac{f_{\max}}{f_{\min}}} & \text{if } \frac{f_{\max}}{f_{\min}} > a, \frac{f_{\min}}{f_{\max}} > b \text{ and } p_m < 1 - b \\ p_m & \text{otherwise} \end{cases} \quad (9)$$

其中, f_{\max} 、 f_{\min} 、 f_{ave} 分别表示种群的最大适应度、最小适应度以及平均适应度, f_{\max} 与 f_{\min} 的接近程度反映了整个种群的集中程度; f_{ave} 与 f_{\max} 的接近程度反映了种群内部适应度的分布情况。参数 $0.5 < a < 1$, $0 < b < 1$, 当 $f_{ave}/f_{\max} > a$ 且 $f_{\min}/f_{\max} > b$ 时, 判断为种群“集中”, 使 p_c 和 p_m 根据种群的集中程度自适应变化; 不满足该条件时, 判断为种群“分散”, p_c 和 p_m 保持最初的较小的初值。文中, 取 $a = 0.65$, $b = 0.01$ 。

2.5 保留最优路径

把 MA 路径中的每一代遗传操作后产生的新一代种群的最高适应度与上一代种群的最高适应度作比较, 如果小于上一代最高适应度, 就随机淘汰新一代中的一个个体, 把上一代中具有最高适应度的个体加入到新一代中, 保证当前的最优个体不会被交叉、变异等遗传运算破坏。

2.6 终止判据

算法若满足以下任何一个条件: 算法达到设定的最大进化代数 GEN_MAX; 适应值的最大值和最小值的相对差值小于 10^{-6} , 则输出当前节点访问序列 P 为最优路径, 算法结束。

3 算法分析与仿真

3.1 算法分析

根据对 DAGGMA 算法的描述可知, 算法的各个

阶段是顺序进行的。设 WSN 由 n 个节点组成。在网格建立过程中, 算法只需要一个 for 循环来实现, 而且每个节点最多发送一个 HELLO 报文即可, 所以其时间复杂度 (Time Complexity, TC) 和消息复杂度 (Message Complexity, MC) 均为 $O(n)$ 。在其他阶段中, 算法的复杂度主要由 AGA 的适应度函数的设计和编码策略决定, 其 TC 为 $O(nPG)$, 其中 P 是初始群体中个体的数目 (即 WSN 中网格的数目), G 指最大进化代数 GEN_MAX, 其 MC 为 $O(n)$ 。

3.2 仿真实验与分析

我们采用 NS 仿真器作为实验平台, 模拟实验是在如下环境下进行的: 网络部署在 $200\text{ m} \times 200\text{ m}$ 矩形区域中, 在网络随机散布 N 个节点, N 从 50 变化到 300 (以 50 为变化单位), 节点的通信半径为 R , Sink 节点建立边长为 $r \times r$ 的网格 (其中 $r = R/2.5$), Sink 节点固定, 位于网络的左下角。传感节点最大传输距离由网络部署区域面积和节点数确定, 若节点间以 95% 的概率连通, 则节点间最大传输距离应该满足 $R_{\max} = 200 \times 200 / \sqrt{0.224535N + 0.657175}^{[11]}$ 。目标节点的发射能量为 1000 单位, 相对传感节点感知能力, 取目标信号最大传播距离 $D_{\max} = R_{\max}$ 。假定移动代理本身代码长度为 100 byte, 携带的数据大小为 400 byte。最大迭代数 GEN_MAX = 200。仿真采用的 MA 模型分别是 LCF 算法与本文提出的 DAGGMA 算法。算法的性能主要从网络能耗 (能量消耗) 与网络延时两个方面进行评估。

3.2.1 网络能耗分析

在基于 MA 的 WSN 中, 网络中能耗分为两大部分: 一是为 MA 规划路由的能耗, MA 规划路由的能耗评估路径代价反映了路径的资源消耗; 二是 MA 在 WSN 中迁移并在目标节点进行数据融合所消耗的能量, 即优化路径的信息增益的能量消耗。算法优化路径的信息增益必须满足应用所需的最小门限, 所以用 MA 规划路由的能耗与路径总信息增益的能耗的比值来考察单位信息增益所付出的能量消耗代价。设节点的通信半径 $R = 50\text{ m}$, 则网格的边长 $r = 20\text{ m}$, 图 2 给出了在不同节点规模下单位

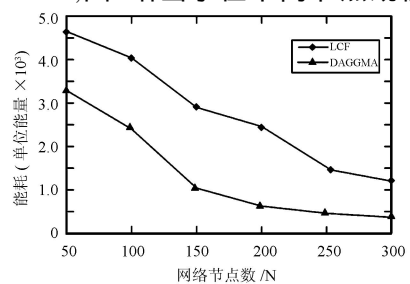


图 2 两种算法模式下单位信息量的网络能耗比较

信息增益的能量消耗代价的关系。由图可见,随着节点规模的增长,平均信息增益能耗呈下降趋势,体现了 WSN 发展中“以数量换质量”的特点。LCF 为局部搜索方法,其性能比 DAGGMA 要差,这是因为 LCF 算法中,移动代理在每个节点进行数据融合,没考虑均衡所有节点的能量消耗;而在 DAGGMA 算法下,只有处于移动代理沿最优路径上的网格中的节点才进行数据融合,因而相同条件下 DAGGMA 算法可维持更长的网络寿命。

3.2.2 网络延时分析

在基于 MA 的 WSN 中,网络延时是指从 Sink 节点准备分派移动代理开始到移动代理携带数据融合结果返回 Sink 节点所需的时间。设在该网络的带宽下,传送 1 bit 数据需 1 个单位时间。随着目标节点增多时两种算法下的延时如图 3 所示。图 3 表明了 DAGGMA 算法执行数据融合所带来的延时相比于 LCF 算法要小得多,并且基本上不受目标节点规模增长的影响。这是因为在 LCF 算法中,MA 在每个节点进行数据融合,增加了数据处理的时间;而在 DAGGMA 算法中,移动代理沿最优路径访问目标周围的节点,采用量化编码的办法减少了数据迁移和处理过程中的数据量,进行数据融合后沿最优路径返回处理节点,降低了网络延时。

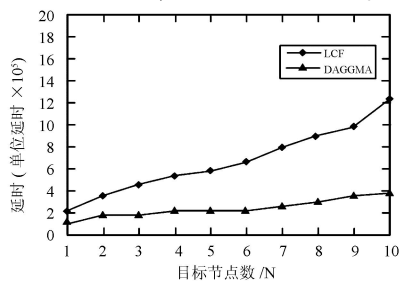


图 3 两种算法模式下网络延时比较

4 结束语

在 WSN 中,针对传感器节点能量有限及传感器网络内节点密度很高的特点,通过适当的数据融合算法可以实现网络节能,延长整个网络的生存时间。本文提出了一种基于网格和移动代理的数据融合算法 DAGGMA。基于移动代理对 Sink 节点发出兴趣代理报文和目标节点发出数据代理报文进行转发,使得移动代理在传感节点间迁移的同时能够进行有效的数据融合。在 DAGGMA 算法中,将移动代理路由归结为一个优化问题,通过把 WSN 均匀分割为多个大小适当的二维网格,形成 AGA 的初始群体,采用 AGA 求出移动代理的最优路由由节点序列。仿真结果表明,随着网络规模增大,和

LCF 算法相比,DAGGMA 算法有更小的网络能耗和延时,延长了网络的生存时间。

参考文献:

- [1] Li Zhiyu, Shi Haoshan. Design of Gradient and Node Remaining Energy Constrained Directed Diffusion Routing for WSN [C]// The 3th International conference on Wireless Communications, Networking and Mobile Computing, Sept. 2007: 2600-2603.
- [2] Boulis A, Ganeriwal S, Srivastava M B. Aggregation in Sensor Networks: an Energy-Accuracy Trade-Off [J/OL]. In Elsevier journal of Ad Hoc Networks, 2003, 1(1): 317-331.
- [3] 周四望等. 无线传感器网络中基于数据融合的移动代理曲线动态路由算法研究[J]. 计算机学报, 2007, 30(6): 894-904.
- [4] Qi H, et al. Multiresolution Data Integration Using Mobile Agents in Distributed Sensor Networks [J/OL]. IEEE Transaction on Systems, Man, and Cybernetics Part C: Applications and Reviews, 2001, 31(3): 383-291.
- [5] Rajagopalan R, et al. Multi-Objective Mobile Agent Routing in Wireless Sensor Networks [J/OL]. IEEE Trans. On congress on Evolutionary Computation, 2005, 5(5): 1730-1737.
- [6] Nen-Chung Wang, et al. Energy-Aware Data Aggregation for Grid-Based Wireless Sensor Networks with a Mobile Sink [J/OL]. Wireless Personal Communications, Springer Netherlands, July 11, 2007: 1-13.
- [7] Uttara Sawant, B. E. Grid-based Coordinated Routing in Wireless Sensor Networks [D]. Master's Thesis, University of North Texas, USA, 2006.
- [8] 陈治平,彭静. 基于网格和曲线转发的传感器网络路由算法[J]. 计算机应用, 2006, 26(4): 914-917.
- [9] 杨少军,史浩山,黄睿. 无线传感器网络移动 Agent 路由算法的研究与仿真[J]. 系统仿真学报, 2007, 19(2): 388-392.
- [10] 陈明杰,刘胜. 改进自适应遗传算法在函数优化中的应用研究[J]. 哈尔滨工程大学学报, 2007, 28(8): 975-879.
- [11] Koskinen H. Connectivity and Reliability in Ad Hoc Networks [D]. Master's thesis, Helsinki University of Technology, Helsinki, Finland, 2003.



李志宇(1977-),男,博士研究生,讲师。主要研究方向:计算机控制与智能控制、无线传感器网络,lizhiyu@nwpu.edu.cn



史浩山(1946-),男,博士生导师,主要研究领域:多媒体通信与网络安全,无线传感器网络,shilaooshi@nwpu.edu.cn