# 南京邮电大学学术型硕士研究生学位论文开题报告

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 学号 | 1016041018 | 姓名 | 郑文添 | 手机 | 18362971301 |
| 专业 | 信息安全 | | 所在学院 | 计算机学院、软件学院 | |
| 已获得的  课程学分 | 29 | | 是否达到  培养计划要求 | 是 | |
| 未完成的课程  及预计完成时间 | 无 | | | | |
| 补修课程及成绩 | 无 | | | | |
| 初定论文题目 | 无线传感器网络异常数据检测 | | | | |
| 论文选题来源 | 学校自选项目 | | | | |
| 论文类型 | 基础研究 | | | | |
| 导师对所选课题的基本要求  查阅相关文献，了解本课题的背景知识和国内外目前的研究进展。对与本课题密切相关的文献进行重点学习研究，总结相关的解决思路和算法过程。试图在当前的理论和方法的基础上，结合发展现状，找到突破点，进行改进或创新。  严格按照进度计划开展课题研究，逐步细化各步骤，及时总结思路，加强与导师、同行间的学术交流，在此基础上，严格遵守硕士学位论文的要求，撰写毕业论文。 | | | | | |

1. 本报告A4纸双面打印，研究生院、学院各一份，导师、学生自留底稿。

|  |
| --- |
| 一、选题依据（综述报告）  （与选题有关的国内外研究综述，阐述研究目的和实际意义，列出主要阅读参考文献，不少于20篇）  （一）与选题有关的国内外研究综述  (1) 无线传感器网络技术  无线传感器网络(Wireless Sensor Network,WSN)是一种新兴的科学技术网络，最早的研究来源于美国军方。它是由部署在监测区域内大量廉价、低功耗、能在微小体积内进行信息采集、数据处理和无线通信等多种功能的微型传感器节点组成，通过无线通信方式形成一个多跳的自组织的网络系统[1]。WSN综合了传感器技术、嵌入式计算技术、分布式信息处理技术和网络通信技术，能够协作地实时监测、感知和采集网络分布区域内的各种环境和监测对象信息，并对这些信息进行初步处理，获得详尽而准确的信息，然后传送给终端用户，以实现复杂的指定范围内目标检测与跟踪。  ①无线传感器网络的体系架构  无线传感器网络的体系架构中主要包括传感器节点(Sensor node)、汇聚节点(Sink node)、和管理节点[2]。传感器节点通过人工布置或者飞行器撒播等方式散布在监测区域内部或附近，这些节点通过自组织方式构成无线网络。传感器节点监测的数据沿着其他传感器节点逐跳地进行传输，在传输过程中监测数据可能被多个节点处理，经过多跳后到汇聚节点，最后通过互联网或卫星到达管理节点。用户通过管理节点对传感器网络进行配置和管理,发布监测任务。传感器节点通常是一个微型的嵌入式系统,它的处理能力、存储能力和通信能力相对较弱,通过携带能量有限的电池供电[3]。汇聚节点的处理能力、存储能力和通信能力相对比较强，它是一个特殊的节点，连接传感器网络与Internet等外部网络。  ②无线传感器网络的特征  a、自组织、无中心的网络  b、动态的网络拓扑型结构  c、网络规模大、节点数量多、密度大  d、传感器节点的节点能量、计算能力和存储能力有限  e、通信能力有限，数据多跳传输  ③无线传感器网络的应用  无线传感器网络的应用前景非常广阔,能够广泛的应用于军事、环境监测和预报、健康护理、智能家居、建筑物状态的监控、复杂机械监控、空间探索和仓库管理,以及机场、大型工业园区的安全监测领域。  ④无线传感器网络的安全需求  无线传感器网络作为任务型的网络，需要进行数据传输、采集和融合以及任务的协同控制等。如何保证任务执行的机密性、数据产生的可靠性、数据融合的高效性和数据传输的安全性，就成为无线传感器网络安全问题需要全面考虑的内容。总的来说,无线传感器网络的安全,需要解决如下问题：点到点认证问题、保密性问题、完整性问题、新鲜性问题、扩展性问题、鲁棒性问题、可用性问题和安全管理问题。  (2) 无线传感器网络异常检测概述  无线传感器网络节点一般部署在无人监守的野外环境中，传感器节点极易受到攻击或者损害，造成传感器节点的硬件出错或者软件出错。硬件出错意味着传感器节点将无法收集数据和发送数据，软件出错意味着传感器内部程序有问题，虽然可以收集数据和发送数据，但是他们所监测的数据往往是错误数据[4](当传感器节点能量很低的时候，检测的数据为异常数据的概率特别高)。数据的可靠性是无线传感器网络安全的重要需求，需要通过异常数据检测机制，将异常数据剔除，保证数据集的可靠性。  ①无线传感器网络中的异常分类  无线传感器网络中的异常定义为与已知数据或者传感器行为截然不同的情况，主要分为行为异常和数据异常。行为异常指的是传感器节点的行为与正常行为不同，例如文献[5,6]中通过传感器节点的速度行为异常检测恶意的副本节点。数据异常指的是传感器节点的监测值和实际值存在较大偏差的数据，分为系统度量值异常和读数异常两种类型。系统度量通常指节点自身的状态或行为数据，如单位时间内与邻居间的数据交换次数、单位时间内接收的数据包和发送的数据包数量等，不同类型间度量之间存在或多或少的关联性，文献[7]中通过研究节点系统度量之间的相关性，有效的检测出异常节点。传感器网络中读数是指温度、湿度、光亮度等传感器节点采集的数据，这些数据的特点是单一类型数据（如温度）与数据采集的时间、地点存在空时关联性，但不同类型数据之间（如温度与湿度之间）不存在显式相关性。文献[8]中通过节点读数的时间相关性和邻居节点读数的空间相关性建立贝叶斯分类器，对异常数据和正常数据进行分类，得到较高的检测率。  ②异常数据的检测方法  异常检测的方法可以分为两大类：集中式检测方案和分布式检测方案[1,4,9].集中式方案中传感器节点通过邻居节点将数据逐跳地传送到汇聚节点进行处理,汇聚节点使用相应的检测模型进行异常检测[10-11]。集中式方案的缺点是:汇聚节点的处理能力要很强，一旦汇聚节点被攻击，那么整个系统就无法进行检测，另外集中式检测将传感器节点的数据全部传送到汇聚节点，大大地增加了网络的通信开销，缩短了网络的生命周期。分布式检测方案中节点通过邻居节点的数据信息或者自身数据的历史信息，实现节点的自动检测[12-14]。分布式方案实现每一个节点的自动检测，减小了汇聚节点的工作量，同时大大减小了网络的通信开销，延长了网络的生存周期。目前的研究方案中更多采用分布式的检测方案，进一步可以将分布式异常检测方案分为五大类型：基于统计的分布检测方案、基于最近邻居的分布式检测方案、基于聚类的分布式检测方案、基于分类器的检测方案、基于频谱分解的检测方案[15]。  (I)基于统计的分布式的检测方案  基于统计的检测方案是最早提出的异常检测方案，这种方案实质上属于基于模型的方案。该方法通过假设或者估计，建立数据满足的分布模型。当有新的数据到达时，比较该数据和模型的匹配程度来判别数据是否异常[16]。如果能够建立正确的数据概率模型，那么基于统计的检测方案在能够很好的检测出异常数据(这个可以通过数学关系式证明)。然而在很多现实场景中，传感器数据分布的先验知识很难获取，这也使得建立一个正确的数据分布模型比较困难。  (II)基于最近邻居的检测方案  基于最近邻居的检测方案是数据挖掘和机器学习领域中相对于其最近邻居分析数据实例的最常用方法，该方案中使用预先设定的距离概念，计算两个数据实例之间的相似性[17,18]。如果一个数据实例和它的邻居节点距离超过某个阈值(两个节点相似性低)，则认为该数据实例是一个异常值。通常在单变量中使用欧式距离衡量两个数据实例的相似性，在多变量中使用马氏距离衡量数据实例相似性。基于最近邻居的检测方案不需要对数据的分布作出任何假设，能够实现基于统计的分布式检测方案的大部分功能。但是该方法需要合适的输入参数，且当计算多变量之间的距离时，计算量较大，不适合在大规模网络中进行检测。  (III)基于聚类的检测方案  基于聚类的检测方案是将相似的数据实例归为具有相似行为的簇中，是数据挖掘中数据分类的重要方法。该方案中认为不属于任何簇群的数据实例或者属于簇群很小的数据实例为异常数据，K均值聚类法是该方法的典型应用[19]。基于聚类的检测方案，不需要知道数据的先验概率分布，并且能够在增量模型中使用，即可以将新的数据实例馈送到系统中并且被检测以找到异常值。然而该方案需要设置合适的簇群大小和簇群个数，并且当计算多元数据实例之间的距离时，需要消耗较大的计算资源。  (IV)基于分类的检测方案  分类器方法是数据挖掘和机器学习领域中一种非常重要的系统方法。该方案通过训练数据集得到一个分类器，然后将测试数据输入该分类器，对数据进行分类(正常数据/异常数据)。非监督的分类器方法不需要有标签的数据集就能训练出良好的分类器，而监督的分类器方法需要标签数据集来训练分类器。常用的分类方法主要有两种：基于支持向量机(SVN)的分类方法[20-22]和基于贝叶斯网络的分类器[23-24]。分类器方法能够很好的实现数据的分类，具有很高的数据检测率。然而基于支持向量机的分类方法会产生较大的计算开销，对传感器节点资源消耗较大，且需要选择合适的内核函数。基于贝叶斯网络的分类法受到传感网中变量个数的影响。  (V)基于频谱分解的检测方案  基于频谱分解的检测方案主要是通过分析数据的主成分(PC)来检测数据的正确性。主成分分析法通过减少数据的维度来描述数据的主要特性，主要通过前几个主成分来描述数据的结构，任何与该结构不符合的数据实例被认为是异常数据[25-27]。主成分分析法通过较低维度的数据描述原有的高维数据，可以起到数据降维的作用，但是选择合适的主成分个数描述数据需要较高的计算开销，对传感器资源消耗较大。   1. 研究目的及实际意义   随着近几年物联网技术的飞速发展，无线传感器网络技术受到越来越多的关注，特别是在军事侦  查和环境检测等运用中，在数据传输和监测过程中如何保证数据的准确性显得尤为重要。通常情况下，无线传感器网络测量和收集的数据是不可信的，数据的准确性会受到诸如噪声、错误、数据丢失、重复数据等因素的影响。无线传感器节点价格低廉、节点能量有限、计算资源有限、存储资源有限、通信带宽有限，这些资源限制同样造成传感器节点测量数据不精确。另一方面，传感器节点的运作受到环境因素的影响，将传感器节点布置在环境恶劣，无人值守的野外环境下，传感器节点可能会运行异常，导致数据的出错。在现实世界中，诸如森林火灾，地震预警等应用需要非常准确的数据来提供预测，因此必须保证传感器数据的可靠性和准确性。异常数据是影响数据质量的重要因素，通过异常数据检测模型将异常数据从数据集中剔除，保证数据集的准确性，是研究的主要目的。前人已经在这方面做了很多工作，但是依然存在很多缺陷，如需要消耗大量的计算资源、引入了巨大的通信开销、检测精度不高，误报率高等问题。我们需要对前人的工作进行总结，找到适合于无线传感器网络的异常检测模型即：高准确率、低误报率、不会给网络带来巨大的资源消耗的方案。同时在前人的工作基础上，我们需要进一步分析异常数据产生的原因，是单纯的节点数据错误还是有恶意的事件发生，或者传感器节点遭受了攻击被捕获。   1. 参考文献 2. McDonald, D., et al. "A Survey of Methods for Finding Outliers in Wireless Sensor Networks." Journal of Network and Systems Management.2015, 23(1): 163-182. 3. 孙利民，李建中，陈渝.无线传感器网络.北京:清华大学出版社,2005,4-14 4. Ghorbel, O., et al. One class outlier detection method in wireless sensor networks: Comparative study. 24th International Conference on Software, Telecommunications and Computer Networks (SoftCOM), 2016. 5. Feng, H., et al. "Distributed outlier detection algorithm based on credibility feedback in wireless sensor networks." IET Communications.2017,11(8): 1291-1296. 6. Jun-Won, H., et al. Fast Detection of Replica Node Attacks in Mobile Sensor Networks Using Sequential Analysis. 28th Annual IEEE International Conference on Computer Communcations (INFOCOM),2009. 7. Jun-Won, H., et al. "Fast Detection of Mobile Replica Node Attacks in Wireless Sensor Networks Using Sequential Hypothesis Testing." IEEE Transactions on Mobile Computing (TMC),2011, 10(6): 767-782. 8. Liu, Q., et al. A metric-correlation-based distributed fault detection approach in wireless sensor networks. 17th Asia-Pacific Network Operations and Management Symposium (APNOMS)，2015. 9. Titouna, C., et al. "Outlier Detection Approach Using Bayes Classifiers in Wireless Sensor Networks." Wireless Personal Communications,2015, 85(3): 1009-1023. 10. Zhang, Y., et al. "Outlier Detection Techniques for Wireless Sensor Networks: A Survey." IEEE Communications Surveys & Tutorials,2010, 12(2): 159-170. 11. Bill, C.P.L., Eden, W.M.M., Tommy, W.S.C.: ‘Probabilistic fault detector for Wireless Sensor Network’, Expert Syst. Appl., 2014, 41, (8), pp. 3703–3711 12. Gil, P., Martins, H., Januário, F.: ‘Detection and accommodation of outliers in wireless sensor networks within a multi-agent framework’, Appl. Soft Compute, 2016, 42, pp. 204–214 13. Alessandra, D.P., Salvatore, G., Giuseppe, L.R., et al.: ‘Adaptive distributed outlier detection for WSNs’, IEEE Trans. Cybern., 2015, 45, (5), pp. 888–899 14. Panda, M., Khilar, P.M.: ‘Distributed Byzantine fault detection technique in wireless sensor networks based on hypothesis testing’, Comput. Electr. Eng.,2015, 48, pp. 270–285 15. Panda, M., Khilar, P.M.: ‘Distributed self fault diagnosis algorithm for large scale wireless sensor networks using modified three sigma edit test’, Ad Hoc Netw., 2015, 25, (Part A), pp. 170–184 16. Andrade, A. T. C., et al. Outlier detection using k-means clustering and lightweight methods for Wireless Sensor Networks. IECON 42nd Annual Conference of the IEEE Industrial Electronics Society.2016 17. M. A. Rassam, A. Zainal, and M. A. Maarof, "Advancements of data anomaly detection research in Wireless Sensor Networks: A survey and open issues," 2013. 18. S. Ramaswamy, R. Rastogi, and K. Shim, Efficient Algorithms for Mining Outliers from Large Data Sets, ACM Special Interest Group on Management of Data, pp. 427-438, 2000. 19. E. Knorr and R. Ng, Algorithms for Mining Distance-Based Outliers in Large Data Sets, International Journal of Very Large Data Bases, pp. 392-403, 1998 20. Andrade, A. T. C., et al. (2016). Outlier detection using k-means clustering and lightweight methods for Wireless Sensor Networks. IECON 42nd Annual Conference of the IEEE Industrial Electronics Society.2016 21. Zhang, Y., et al. "Distributed online outlier detection in wireless sensor networks using ellipsoidal support vector machine." Ad Hoc Networks, 2013, 11(3): 1062-1074. 22. Shahid, N., et al. Quarter-Sphere SVM: Attribute and Spatio-Temporal correlations based Outlier &amp; Event Detection in wireless sensor networks. 2012 IEEE Wireless Communications and Networking Conference (WCNC).2012 23. Rajasegarar, S., et al. "Centered Hyperspherical and Hyperellipsoidal One-Class Support Vector Machines for Anomaly Detection in Sensor Networks." IEEE Transactions on Information Forensics and Security (TIFS) .2010, 5(3): 518-533. 24. Hao, Y., et al. A Distributed Bayesian Algorithm for data fault detection in wireless sensor networks. 2015 International Conference on Information Networking (ICOIN).2015 25. Titouna, C., et al. "Outlier Detection Approach Using Bayes Classifiers in Wireless Sensor Networks." Wireless Personal Communications.2015, 85(3): 1009-1023. 26. Ghorbel, O., et al. One class outlier detection method in wireless sensor networks: Comparative study. 2016 24th International Conference on Software, Telecommunications and Computer Networks (SoftCOM).2016 27. Ghorbel, O., et al. A Novel Outlier Detection Model Based on One Class Principal Component Classifier in Wireless Sensor Networks. 2015 IEEE 29th International Conference on Advanced Information Networking and Applications.2015 28. Fernandes Jr, G., et al. "Autonomous profile-based anomaly detection system using principal component analysis and flow analysis." Applied Soft Computing.2015, 34: 513-525. |
| 二、选题的研究目标、研究内容、所要解决的主要问题及可能的创新点  （一）研究目标  分析无线传感器网络的特性，结合具体的网络结构(例如:大规模密集分布、小规模离散分布)，分析异常数据的检测。主要通过节点之间的空域相关性(邻居节点之间数据的相关性)和时域相关性(节点历史数据记录之间的相关性)建立数学模型，对数据进行检测，将异常数据从数据集中剔除，保证数据集的可靠性。常见的检测模型有基于贝叶斯网络的分类器模型，基于支持向量机(SVM)的分类器模型，基于主成分分析法(PCA)的特征检测模型，对上述模型进行改进，提出更加适合于无线传感器网络的异常检测模型(即达到：高检测率、低误报率、资源消耗小等特点)。在此基础上，我们需要提出一种有效的判别机制，识别异常数据发生的原因(单纯的错误数据，恶意事件、恶意攻击)。   1. 研究内容   无线传感器网络区别于一般的无线网络，拥有自身的特性：自组织、无结构、能量有限、计算资源有限、通信带宽有限。这些特点决定了其他领域中的异常检测方法很难应用到无线传感器网络中的异常数据的检测中。本文着重研究无线传感器网络异常数据的检测模型以及异常数据识别机制，主要研究内容如下：   1. 从整体上研究无线传感器网络的特性，以及异常检测的理论与方法，基本掌握异常检测的实现过程。 2. 研究异常检测模型的优化。通过相关理论与阅读文献，找出传统异常检测方法的优点不足，在已有的异常检测模型基础之上进行系统的改进与优化，提出自己的异常检测模型。 3. 研究异常数据发生的原因，利用节点之间的空时相关性，提出有效的异常数据识别机制，识别导致异常数据发生的原因 4. 通过Matlab建立完整的异常检测模型，印证既定的检测方案，取得实验数据，分析检测方案的优劣性，再根据异常数据识别机制识别异常发生的原因。   （三）所要解决的主要问题及可能的创新点  1、所要解决的主要问题：  无线传感器网络区别于一般的无线网络，拥有自身的特性：自组织、无结构、能量有限、计算资源有限、通信带宽有限。这些特点决定了其他领域中的异常检测方法很难应用到无线传感器网络中的异常数据的检测中。如何通过建立合适的异常检测模型，在保证检测率与误报率的前提上减小网络的开销。如今还存在以下一些问题：  （1）传统的分类器识别模型中，通过对数据集进行训练能够得到较为精确的分类器，能够得到较高的检测率，但是对数据训练的过程中，计算过于复杂，消耗了传感器节点的大量资源。  （2）传统的空时相关模型中，能够保证数据的检测率较高，误报率较低，但是引入了空间相关性，会引起额外的通信开销，缩短传感器网络的生存时间。  2、可能的创新点：  创新点1：对前人的研究方法进行整合改进，在保证计算度不会过于复杂的前提下，实现比目前检测方案中高的检测率。  创新点2：在保证计算资源和通信资源较低的情况下，实现比目前检测方案中低的误报率。  创新点3：提出异常数据源检测机制，检测发生异常的原因 |
| 三、研究方法（预期思路或技术路线）及可行性分析  （一）研究方法（预期思路或技术路线）  本课题将通过分析已有的异常数据检测模型，并结合机器学习的相关算法（如：PCA，SVM，EBP）对已有模型进行改进，构建适合于无线传感器网络的异常检测模型。模型主要依据无线传感器网络的两大特性来构建：一是无线传感器节点之间空间相关性，即邻居节点之间的数据存在很强的相关性。二是无线传感器网络节点的时域相关性，即同一个节点的历史数据之间存在很强的相关性。通过理论和实验分析模型的可靠性和高效性。同时，将密切关注跟踪国内外动向，积极与同行交流，获得最新的研究动态和研究成果，促进自己对该问题的研究。  前期通过阅读大量的国内外相关文献，研究无线传感器网络的特点，掌握异常检测算法的基本原理。对前人提出的异常数据检测模型进行研究，分析其优点与不足，并对相关模型进行实现和改进，加深对模型建立的理解。通过对前人工作的详细分析，对异常模型中的不足提出解决方案并结合机器学习算法，建立高可靠性，高效率的异常检测模型。  （二）可行性分析  随着物联网技术的发展，越来越多的学者将研究目光投向了无线传感器网络技术，对这方面的研究已经越来越深入了。很多研究者已经提出了针对无线传感器网络的异常检测模型，虽然这些模型都存在着缺陷，但是我们可以从这些模型中学习如何建立异常检测的模型，有助于我们对无线传感器网络异常数据检测这个模块有更深入地了解。同时，随着机器学习算法越来越成熟，应用范围越来越广，将其和传统的异常数据检测机制结合在一起可以获得更好的性能，有助于提高异常数据检测模型的可靠性。 |
| 四、研究基础与条件  本人本科专业为电子信息工程，系统的学习了有关通信原理，移动通信原理，计算机网络等课程，同时参加过全国大学生数学建模比赛，对无线通信和数据建模有一定的基础。研究生阶段，通过专业课程深入学习了无线网络安全方面的知识，对无线传感器网络有了进一步的认识。在课余时间阅读相关文献，对异常数据检测技术进行了认真的归纳与分析，了解异常数据检测机制在无线传感器网络中的重要性，因此初步确定此研究方向。计划每周阅读一些相关资料，以此为研究本课题打下良好的基础，积累大量的经验和技术依据。因此，本课题研究的理论基础、技术基础已基本具备，加之有丰富研究经验的导师带领指导，基本能顺利完成本课题的研究，取得预期的研究成果。 |
| 五、研究进度及具体时间安排（包括起讫日期、主要研究内容和预期结果）  2017.8-2017.10 进行课题调研，筛选、查阅有关文献资料，对相关内容进行总结和整理，完成开题报告；  2017.11-2018.05 深入研究相关文献，进一步学习相关理论，整理和分析课题需求，总结当前异常检测机制和方法的优点及不足，熟悉所需要的关键技术并提出疑难问题，及时解决；  2018.06-2018.11 确定具体的研究点和创新点，拟定异常数据检测模型，提出自己的想法，并着手实践，进行仿真实验和结果分析，完成相关成果；  2018.12-2019.01 就前期工作内容中出现的问题及解决方案进行总结，拟定毕业论文的框架和内容，撰写学位论文；  2019.02-2019.03 学位论文评阅，准备论文答辩。 |
| 导师对开题报告的意见  签名： 日期： 年 月 日 |