

**UNIVERSIDADE ESTADUAL DE PIAUÍ
CAMPUS POETA TORQUATO NETO**

Danilo Falcão Barrinha

**REDUÇÃO DA DIMENSIONALIDADE DE DADOS EM
REDES DE SENSORES SEM FIO UTILIZANDO ACP**

**TERESINA
2018**

DANILO FALÇÃO BARRINHA

REDUÇÃO DA DIMENSIONALIDADE DE DADOS EM
REDES DE SENSORES SEM FIO UTILIZANDO ACP

**Trabalho de Conclusão de Curso sub-
metido à Universidade Estadual do Pi-
auí, como requisito necessário para ob-
tenção do grau de Bacharel em Ciência
da Computação**

Teresina, julho de 2018

UNIVERSIDADE ESTADUAL DO PIAUÍ

DANILO FALÇÃO BARRINHA

Esta Monografia foi julgada adequada para a obtenção do título de Bacharel em
Ciência de Computação, sendo aprovada em sua forma final pela banca examinadora:

Presidente: Prof. Dr. José Martins Bringel
Filho
Universidade Estadual do Piauí - UESPI

Membro: Prof. José de Anchieta Araujo
Marques
Universidade Estadual do Piauí - UESPI

Membro: Prof. Herbert Meneses dos Santos
Universidade Estadual do Piauí - UESPI

Teresina, 7 de janeiro de 2019

Agradecimentos

Primeiramente agradecer a Deus por tudo. Meus pais por todo o apoio e dedicação para me proporcionar um ensino de qualidade, além da formação de meu caráter. Meus amigos sempre presentes, não só nos momentos alegres, mas também nos de dificuldade. Agradeço à Universidade Estadual do Piauí por todo o suporte para nossa formação. Aos professores do curso de Ciência da computação em especial ao professor Dr. Marcos Vinicius de Sousa Lemos pela orientação, apoio e confiança.

“Benkyou, Benkyou, Benkyou...”
(Oe Kintaro)

Resumo

As Redes de sensores sem fio são compostas por um conjunto de dispositivos, denominados nós sensores, capazes de sensoriar um determinado fenômeno e transmitir, por meio de um canal sem fio, as informações para uma estação-base. Estes nós sensores possuem uma série de restrições, como capacidade de processamento, memória, e energia, que impactam no desenvolvimento de aplicações para este tipo de rede. Considerando que os nós em muitos casos são operados por bateria e que a transmissão de dados pelo meio sem fio é responsável pela maior parte do consumo de energia, faz-se necessário o desenvolvimento de algoritmos e protocolos que otimizem o número de pacotes que trafegam pela rede. A principal abordagem encontrada na literatura é a redução de dados, dentre as técnicas, pode ser citado a utilização de modelos baseados em séries temporais, que consistem na criação de modelos de predição que representam um conjunto de dados (redundantes ou correlacionados temporalmente) que não são transmitidos pelos nós. Entretanto, tais modelos geralmente são baseados em apenas uma variável/dimensão, de forma que é necessária a criação de vários modelos quando o nós faz o sensoriamento de múltiplas condições ambientes. Dessa forma, neste projeto foi desenvolvido um algoritmo denominado PCASW que utiliza do método de Análise de Componentes Principais (ACP) tomando como base o funcionamento do algoritmo de janela deslizante utilizado pelo protocolo TCP/IP. Após os testes, foi notado uma redução de aproximadamente $1/n$ nos dados utilizados no processo, sendo n o número de dimensões participantes.

Palavras-chave: Rede de Sensores Sem Fio; Análise de componentes principais; redução de dados.

Abstract

Wireless sensor networks are composed of a set of devices, called sensor nodes, capable of sensing a certain phenomenon and transmit, through a wireless channel, the information to a base station. These sensor nodes have a number of constraints, such as processing power, memory, and power, that impact the development of applications for this type of network. Considering that the nodes in many cases are battery operated and that wireless data transmission is responsible for most of the energy consumption, it is necessary to develop algorithms and protocols that optimize the number of packages that travel through the network. The main approach found in the literature is the reduction of data, among the techniques, we can mention the use of models based on time series, which consists of the creation of prediction models that represent a set of data (redundant or temporally correlated) that does not are transmitted by the nodes. However, such models are usually based on only one variable / dimension, so it is necessary to create several models when the node senses multiple environments. Thus, in this project an algorithm called PCASW was developed that uses the Principal Component Analysis (PCA) method based on the operation of the sliding window algorithm used by the TCP / IP protocol. After the tests, a reduction of approximately $1 / n$ in the data used in the process, being in the number of participating dimensions.

Keywords: Wireless Sensor Network; Principal Component Analysis; Data Reduction.

Lista de ilustrações

Figura 1 – Exemplo de RSSF típica	16
Figura 2 – Estrutura de um nó sensor	16
Figura 3 – Técnicas de Economia de energia em RSSFs	17
Figura 4 – Redução de dados	18
Figura 5 – Rotação aplicada em conjunto de dados bidimensionais	19
Figura 6 – Exemplo de gráfico de variancia	19
Figura 7 – Algoritmo PCASW	22
Figura 8 – Quadro 802.15.4	24
Figura 9 – Erro médio por medição em cada Nó	26
Figura 10 – Energia gasta com e sem PCASW	27

Lista de tabelas

Tabela 1 – Notações Utilizadas na Explicação da abordagem	21
Tabela 2 – Parâmetros da Simulação	26

Lista de abreviaturas e siglas

ACP - Análise de Componentes Principais

PCA - Principal Component Analysis

RSSF - Redes de Sensores Sem fio

EQM - Erro Quadrático Médio

PCASW - Principal component analysis for sensor network based on sliding window

Sumário

1	INTRODUÇÃO	12
1.1	Objetivo	13
1.1.1	Objetivo Geral	13
1.1.2	Objetivos Específicos	13
1.2	Organização	13
2	FUNDAMENTAÇÃO	15
2.1	Redes de Sensores sem Fio	15
2.1.1	Arquitetura de uma Rede de sensores sem fio	15
2.2	Redução de Dados	17
2.2.1	Análise de Componentes Principais	18
3	ABORDAGEM PROPOSTA	21
3.1	Premissa	21
3.2	PCASW	21
3.2.1	Funcionamento	22
4	AVALIAÇÃO DE PERFORMANCE	23
4.1	Base de dados utilizada	23
4.2	Ferramentas	23
4.3	Métricas de Avaliação	23
4.3.1	Erro Quadrático Médio (EQM)	24
4.3.2	Consumo de Energia	24
5	RESULTADOS	26
5.1	Erro	26
5.2	Energia	27
6	CONCLUSÃO E TRABALHOS FUTUROS	28
6.1	Conclusão	28
6.2	Trabalhos Futuros	28
	REFERÊNCIAS	29

1 Introdução

As redes de sensores sem fio (RSSF) podem ser classificadas como um tipo especial de rede ad hoc de múltiplos saltos [Margi 2017], ou seja, elas não tem um nó central para qual todas as comunicações são direcionadas e as envia para seus respectivos caminhos, na verdade cada nó funciona basicamente como um roteador. Consistem em um grande conjunto de dispositivos contendo sensores e/ou atuadores (nós sensores) capazes de nos fornecer informações sobre algum fenômeno. Elas têm sido aplicadas em diversas áreas, tais como a engenharia ambiental, monitoramento de ambientes, uso industrial, segurança e etc. Devido às características deste tipo de rede, elas possuem restrições de energia, largura de banda e tempo de resposta [AQUINO 2008]. O fato de serem alimentadas por bateria, faz com que energia seja um recurso bastante crítico, pois nem todos os nós sensores podem ser acessados tão facilmente. Existem casos em que os nós se encontram em locais de difícil acesso, que vão desde campos vulcânicos até no monitoramento de ambientes aquáticos. Devido a estas restrições e tendo em vista que a transmissão de informações é o maior responsável pelo gasto de energia, é necessário adotar medidas para a redução no número de transmissões, com o intuito de prolongar a vida útil da rede. Como exemplo, é possível utilizar de algoritmos de computação natural, com o exemplo do protocolo EEABR (Energy-Efficient Ant-Based Routing) que é baseado no algoritmo de colônia de formigas, e busca encontrar caminhos mais eficientes em termos de energia [Camilo et al. 2006]. Outro tipo de solução é a utilização de algoritmos para a redução dos dados a serem transmitidos.

Técnicas para a redução de dados são bastante comuns, uma vez que uma grande quantidade de dados não dificulta somente a transmissão, mas também afeta consideravelmente o tempo de processamento em determinados casos. Como exemplo deste tipo de abordagem, temos os algoritmos baseados em séries temporais, que consistem em observações feitas em um intervalo de tempo, e como resultado é obtido um modelo preditivo baseado nas observações de entrada [Dias, Bellalta e Oechsner 2016], porém a maioria destes métodos supõe somente um tipo de dado, ou seja, é assumido que existe somente um sensor no nó estudado. Dito isto, temos que vários pontos ainda não são explorados se tratando de dados multivariados neste tipo de técnica, com o exemplo da correlação entre tipos de dados. A técnica a ser apresentada neste trabalho explorará este recurso através de métodos de Redução da dimensionalidade de dados multivariados, proporcionando a diminuição dos pacotes a serem enviados. O resultado disto é a redução considerável do número de transmissões, portanto, um menor custo de energia. A técnica abordada neste trabalho é baseada na Análise de componentes principais (ACP), que tem por objetivo prover um novo conjunto de dados menor que os anteriormente coletados, preservando sua

características analíticas. O novo conjunto de dados é chamado de componentes principais, que são nada mais que uma combinação linear das coordenadas originais.

Serão realizados estudos a respeito deste método, e ao final, será um algoritmo que torne eficiente o uso do ACP em RSSF, reduzindo os dados ao máximo com base na taxa de erro máxima informada pelo usuário.

1.1 Objetivo

1.1.1 Objetivo Geral

Desenvolver uma solução que reduza ao máximo o número de pacotes enviados pelo nó sensor, preservando as características analíticas dos dados originais levando em consideração o contexto de dados multidimensionais.

1.1.2 Objetivos Específicos

Considerando o desenvolvimento do trabalho e o objetivo geral apresentado, destacam-se os seguintes objetivos específicos:

- Desenvolver um algoritmo, baseado na técnica Análise de Componentes principais, para reduzir a transmissão de dados em Redes de Sensores Sem Fio objetivando prolongar o tempo de vida destas redes;
- Reduzir o consumo de energia gasta pelos nós sensores em até $1/n$, sendo n o número de dimensões originais;
- Realizar testes comparativos com cenário sem a utilização do algoritmo;
- Análise dos resultados obtidos;

1.2 Organização

O presente trabalho está organizado da seguinte forma:

Capítulo 2: será apresentado todos os conceitos necessários para a compreensão do trabalho. Conceitos relacionados as Redes de sensores sem fio, bem como o levantamento dos principais problemas relacionados à este tipo de rede atualmente, além de conceitos relacionados à redução da dimensionalidade de dados.

Capítulo 3: Será mostrado a abordagem proposta.

Capítulo 4: Avaliação da performance do abordagem proposta, mostrando o cenário e as ferramentas utilizadas.

Capítulo 5: os resultados das simulações realizadas.

Capítulo 6: Conclusão e os Trabalhos futuros.

2 *Fundamentação*

Além das características abordadas anteriormente, outros conceitos são necessários para uma boa compreensão em relação à redução da dimensionalidade de dados em redes de sensores sem fio com o ACP.

2.1 Redes de Sensores sem Fio

Houve um grande avanço nas áreas de sensores, circuitos integrados e comunicação sem fio na última década [Loureiro et al. 2003], proporcionando a criação das RSSFs que conhecemos hoje. Por ser uma área recente e ter certas particularidades que diferem das redes tradicionais, vem tornando-a uma área de grande interesse nos últimos anos. RSSFs são uma subclasse das redes ad hoc, ou seja, cada nó tem um funcionamento similar a um roteador, mas apesar disto existem diferenças das redes ad hoc tradicionais, dentre estas diferenças destacam-se:

- Alta densidade de nós em rede;
- Mudança de topologia muito frequente;
- Comunicação broadcast, diferente da comunicação ponto-a-ponto das tradicionais redes;
- Nós com restrições em termos de energia, processamento e memória ;
- Nós sensores são propensos a falhas;

Este tipo de rede tende a ser autônoma, ou seja, não existe nenhuma, ou quase nenhuma interferência humana, portanto requer uma boa cooperação de cada nó na execução de suas tarefas. Desse modo, os atuais algoritmos distribuídos utilizados nas redes tradicionais, devem ser revistos antes de serem utilizados neste tipo de rede.

RSSFs tem sido aplicadas com sucesso em diversas áreas, tais como, monitoramento de vias e estacionamentos, monitoramento de câmaras frias, monitoramento de ambientes de um prédio, ou até mesmo na agricultura [Bogena et al. 2010].

2.1.1 Arquitetura de uma Rede de sensores sem fio

Uma Rede de sensores sem fio é descrita em Akyldiz et al. [Akyildiz et al. 2002] como um grande conjunto de nós (Figura 1) contendo sensores e/ou atuadores capazes

de nos fornecer informações sobre algum fenômeno (temperatura, humidade, luz) com determinada finalidade. Os dados coletados por cada nó sensor são enviados a um nó central ou sorvedouro, que tem a capacidade de se comunicar com uma estação de controle, onde os dados serão entregues ao seu usuário final. Toda comunicação entre os nós pode ser feita através de múltiplos saltos ou qualquer outro algoritmo definido pelo protocolo de roteamento [Henning, Fonseca e Munaretto].

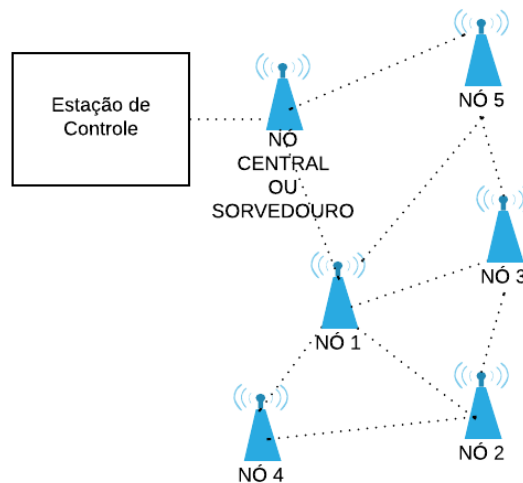


Figura 1 – Exemplo de RSSF típica

Existem deficiências neste tipo de rede que têm sido o foco de vários estudos a fim de resolvê-los, dentre elas temos em destaque neste trabalho, o alto consumo de energia, isto acontece porque na maioria dos casos, estes nós são alimentados por baterias (Figura 2) e se encontram em locais de acesso trabalhoso na maioria dos casos. Isso faz com que eficiência ocupe um dos principais lugares na lista de critérios de projeto de uma RSSF.

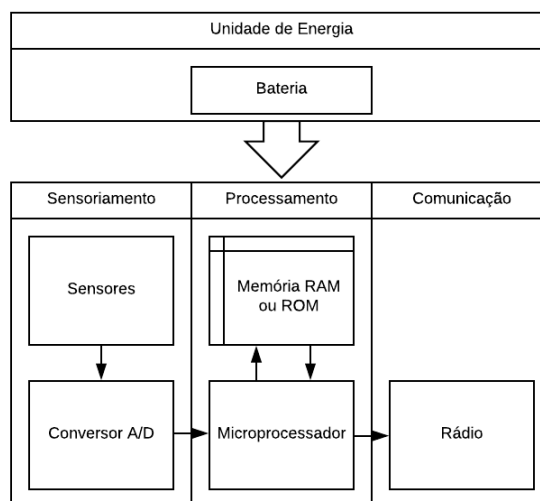


Figura 2 – Estrutura de um nó sensor

Com base nisto, neste trabalho será considera que o gasto em energia para transmis-

são e recepção de informações em um nó sensor serão estimados pelos modelos propostos por [Bezerra et al. 2014] dado pelas equações:

- $Q_{transmissão} = 3 * 21.5mA * (0.26 * 10^{-6}bit/s * TamanhoDados)$
- $Q_{recepção} = 3 * 7.0mA * (0.26 * 10^{-6}bit/s * TamanhoDados)$

Quando:

- $EnergiaDissipada(Q) = Voltagem * CorrenteElétrica * Tempo$
- $Tempo = TaxaDeTransmissão * TamanhoDados$

A taxa de transmissão é considerada como $0.26\mu bits/s$, a corrente elétrica para recepção é considerada $7.0mA$ enquanto a corrente elétrica para a transmissão é fixada em $21.5mA$. Como pode ser observado a partir dos modelos acima, o consumo de energia em uma ação de transmissão é aproximadamente 33% maior que a ação de recepção dos dados, fazendo com que a transmissão de dados seja considerada a causa principal no consumo de energia. Com base nisto, foram desenvolvidas técnicas que tentam mitigar esta deficiência. Já foram desenvolvidos métodos para a redução da quantidade de informações que trafegam pela rede, que vão deste algoritmos para de compressão e fusão de dados, até o uso de técnicas preditivas [Teixeira 2005], como pode ser observado na figura 3.

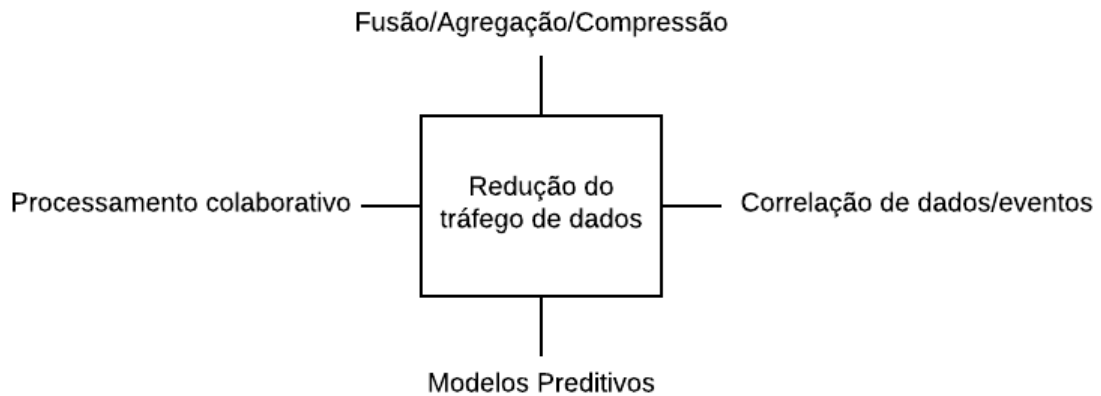


Figura 3 – Técnicas de Economia de energia em RSSFs

2.2 Redução de Dados

Durante a última década diversas áreas das ciências obtiveram um desenvolvimento bastante acelerado, com sistemas avançados de coleta de dados, com isso, o resultado é

um enorme aglomerado de dados. Em muitos casos estes datasets possuem um número tão elevado de variáveis e amostras que acabam tornando a análise bem mais complexa e o processamento bem mais caro computacionalmente se toda informação for considerada. Isto acontece em várias áreas, tais como o processamento de imagens, biomedicina, processamento de linguagem natural, entre outras. O objetivo da redução de dados é chegar a um conjunto de dados menor em volume (Figura 4), mas que produza os mesmos (ou quase os mesmos) resultados analíticos que os dados originais.

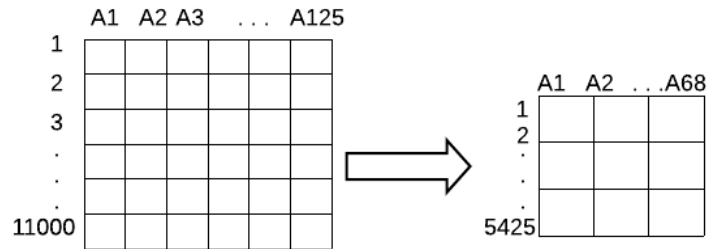


Figura 4 – Redução de dados

Existem diversas técnicas para a redução de dados enviados por um nó sensor, entre eles temos o agrupamento de sensores com base na sua similaridade de dados [Lemos M. e Filho 2017], uma abordagem que visa minimizar o número de nós selecionados para a transmissão de dados, levando a uma diminuição no gasto de energia em até 73.97%. Esta abordagem demonstra um melhor resultado se comparado aos algoritmos de agrupamento baseados na proximidade física. Além destes métodos existem outros, como a Regressão Linear [Carvalho et al. 2011], que tem por objetivo definir modelos lineares, ou seja, retas com o intuito de diminuir o número de transmissões com o mínimo de erro, além desta, uma das técnicas mais usadas para a redução da dimensionalidade atualmente, chamada de Análise de Componentes principais em que a ideia chave é encontrar um novo conjunto de coordenadas em que os dados de entrada podem ser expressos com menos variáveis e com baixo erro.

2.2.1 Análise de Componentes Principais

A ACP, também conhecido como transformada de Hotelling, decomposição ortogonal própria ou transformada de Karhunen-Loève, se trata de uma técnicas que utilizada de princípios matemáticos para transformar um conjunto de números possivelmente correlacionados em um número menor de variáveis. Esta é uma ferramenta bastante popular na redução da dimensionalidade de dados multivariados e eliminação de sobreposições, seu uso mais comum é como o primeiro passo na análise de grandes conjuntos de dados. Outras aplicações mais comuns podem ser:

- Eliminação de ruídos em sinais;

- Compactação de dados.

O processo realizado pelo ACP consiste em uma rotação no plano n-dimensional (Figura 5), que nada mais é do que uma transformação linear, em que a primeira coordenada aponta para a direção de maior variância, a segunda aponta para a segunda direção de maior variância e é ortogonal à anterior, e assim por diante, até a enésima coordenada.

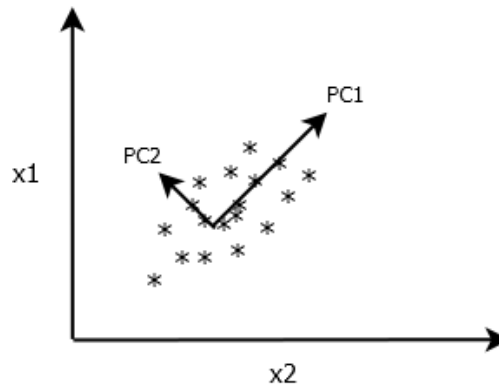


Figura 5 – Rotação aplicada em conjunto de dados bidimensionais

Com isso, temos que a primeira coordenada explica a maior parte da variância total dos dados, já a última componente tem a menor, ou seja, é bem menos informativa, podendo ser descartada (Figura 6). Dessa forma, entre as n componentes principais calculadas, são escolhidas k componentes, onde $k < n$, de tal forma que expliquem a maior variância dos dados, nos possibilitando uma análise com a menor taxa de erro possível, sem perdas significativas. Na figura 6 é mostrando o percentual de variância correspondente a cada componente do dataset utilizado como exemplo.

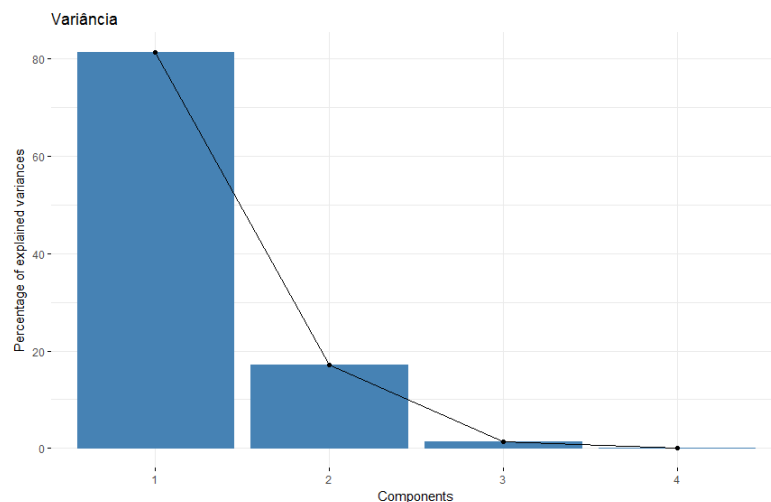


Figura 6 – Exemplo de gráfico de variancia

É possível observar que somente com a primeira componente já temos quase que 82% da explicação total dos dados, quando utilizada as duas primeira componentes, já

temos aproximadamente 93% da explicação total dos dados, podendo ser reconstruído os valores originais com erro quase nulo.

A característica de maior importância encontrada nas componentes principais, é que elas não têm correlação entre si [JACKSON 2003], garantindo assim a não redundância e que seja obtido um novo conjunto de dados com propriedades para análise multivariada, ou seja, ACP transformará um conjunto de dados possivelmente correlacionadas em outro conjunto de variáveis linearmente não correlacionadas.

3 Abordagem Proposta

Neste capítulo o trabalho desenvolvido será descrito, mostrando o foco do algoritmo desenvolvido, os parâmetros do usuário, o processo de validação e o funcionamento da solução proposta. Abaixo na tabela 1, é possível verificar as notações utilizadas para compreensão do trabalho.

Tabela 1 – Notações Utilizadas na Explicação da abordagem

Notação	Descrição
n	Número de dimensões participantes do processo
k	Número de dimensões Resultantes do processo
eqm	Erro Quadrático Médio
t	Tamanho da Janela
tmax	Tamanho máximo da Janela informado pelo usuário
df	Data Frame com medições da janela
dfT	Data Frame temporário com medições sem variância

3.1 Premissa

Este trabalho se concentra em mitigar deficiências em aplicações de monitoramento nas quais os sensores coletam periodicamente informações de determinada área geográfica. Considera-se que os nós sensores são heterogêneos, ou seja, vários tipos de sensores, quanto maior o número de variáveis, maior a redução, uma vez que a equação de redução é dada por $1/n$, sendo n o número de dimensões participantes do processo.

O software Desenvolvido será utilizado em cada nó após cada coleta de dados, com intervalos definidos pelo usuário, fazendo com que cada transmissão de dados seja de forma reduzida.

3.2 PCASW

O algoritmo desenvolvido, denominado PCASW (principal component analysis for sensor network based on sliding window), com o objetivo de diminuir o número de transmissões de um nó sensor, através de técnicas de redução da dimensionalidade de dados. Foi implementado em linguagem R e validado por meio de simulações com datastes de Redes de Sensores reais a partir de modelos de energia propostos por [Bezerra et al. 2014]. O algoritmo trabalhará como janelas, semelhante ao método de controle de fluxo por janela deslizante utilizada pelo TCP/IP, aumentando a janela enquanto o taxa de erro

ainda estiver aceitável, combinado com a técnica da Análise de Componentes Principais, como pode ser observado na Figura 7.

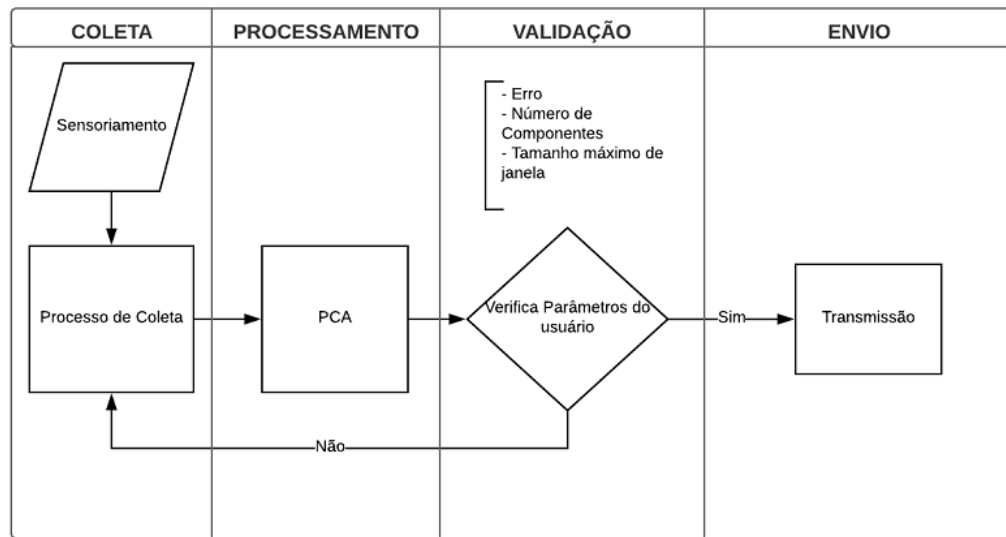


Figura 7 – Algoritmo PCASW

Além disso, é feita uma otimização nas variáveis de cada janela, verificando se a variância é zero ou próximo disso (o que ocorre com mais frequência em sensores de luminosidade para ambientes naturais, as vezes a modificação deste valor demora horas), podendo inferir assim que não há tanta modificação dos dados, podendo ser feita uma média e enviado somente um valor, ao invés de toda a janela, esta funcionalidade é de escolha do usuário.

Existem parâmetros fornecidos pelo usuário, tais como tamanho máximo da janela, fazendo assim com que o usuário não fique tanto tempo sem resposta, além da escolha do erro quadrático médio aceitável.

3.2.1 Funcionamento

O tamanho da janela(t) é iniciado em 10 para melhor resultados na geração do modelo e tem como teto a variável t_{max} (informada pelo usuário), a cada iteração do algoritmo é adicionado uma nova leitura à df . Antes da geração do modelo são feitas algumas verificações, primeiro é verificado a ausência de variância em alguma coluna, se sim, esta coluna é retirada e guardada em uma data frame temporário (dtT) para mais a frente ser verificado a média de valores e consequentemente enviado somente um valor, inclusive existem casos de todas as colunas não terem variância, interrompendo o fluxo do programa e pulando para a próxima janela.

Após isto é verificado se $t \leq t_{max}$ e a geração do modelo ACP, se a explicação dos dados da primeira componente de k é $\geq 90\%$, para que eqm seja mínimo. é feito a transmissão e df é zerado, para dar início à uma nova janela de envio.

4 Avaliação de Performance

Neste capítulo serão apresentados todas as ferramentas utilizadas no processo de simulação da abordagem proposta, bem como os dados utilizados e as métricas de avaliação.

4.1 Base de dados utilizada

Em toda a avaliação de performance deste algoritmo, foi utilizado datasets com medições reais fornecidos por *Intel Lab*. Os dados correspondem a aferições recolhidos a partir de 54 sensores implantados no laboratório Intel Berkeley Research entre 28 de fevereiro e 05 de abril de 2004.

São utilizados sensores do tipo Mica2dot com placas de tempo, as aferições são feitas a cada 31 segundos com informações de tempo, juntamente com a humidade, temperatura, luz e tensão. Os dados foram recolhidos utilizando o TinyDB sistema de processamento de consultas em rede, construído no TinyOS plataforma. São aproximadamente 2,3 milhões de leituras recolhidos por estes sensores. Para as simulações foram consideradas 4 dimensões, temperatura, humidade, luz e energia do nó sensor.

4.2 Ferramentas

Dentre as ferramentas utilizadas para a realização dos testes, foram utilizadas as seguintes:

- **Linguagem R** - é uma linguagem de programação e também um ambiente de desenvolvimento integrado para cálculos estatísticos e gráficos. Ela possui uma série de funções e técnicas já desenvolvidas que foram utilizadas neste trabalho, inclusive o próprio ACP. [wikipedia]

4.3 Métricas de Avaliação

A seguir serão especificados as duas métricas de avaliação utilizados para mensurar a performance do algoritmo comparado ao envio sem a utilização da técnica de redução de dados. A primeira métrica é o Erro quadrático médio associado a cada dataset utilizado nas simulações. Nela será verificado a taxa de erro após a utilização do ACP e comparado às medições originais, a fim de mensurar a eficácia do algoritmo em questão. A outra métrica de avaliação utilizado é a quantidade de energia gasta para o envio de cada dataset.

4.3.1 Erro Quadrático Médio (EQM)

Considerando o conjunto de medições dentro de uma janela transmitida(dfACP), é possível calcular o erro quadrático médio associado em relação às medições originais(dfOrig). Desta forma será possível mensurar o grau de precisão das medições enviadas.

A erro quadrático médio de cada leitura enviado pode ser representado a partir da seguinte equação:

$$EQM = \sqrt{\frac{(\text{dfOrig} - \text{dfACP})^2}{N^{\circ}\text{medições}}}$$

4.3.2 Consumo de Energia

Para a avaliação do gasto de energia gerado por cada sensor foram utilizados os modelos propostos por [Bezerra et al. 2014] dado pelas equações:

- $Q_{\text{transmissão}} = 3 * 21.5mA * (0.26 * 10^{-6} \text{bit/s} * \text{TamanhoDados})$
- $Q_{\text{recepção}} = 3 * 7.0mA * (0.26 * 10^{-6} \text{bit/s} * \text{TamanhoDados})$

Quando:

- $EnergiaDissipada(Q) = \text{Voltagem} * \text{CorrenteElétrica} * \text{Tempo}$
- $\text{Tempo} = \text{TaxaDeTransmissão} * \text{TamanhoDados}$

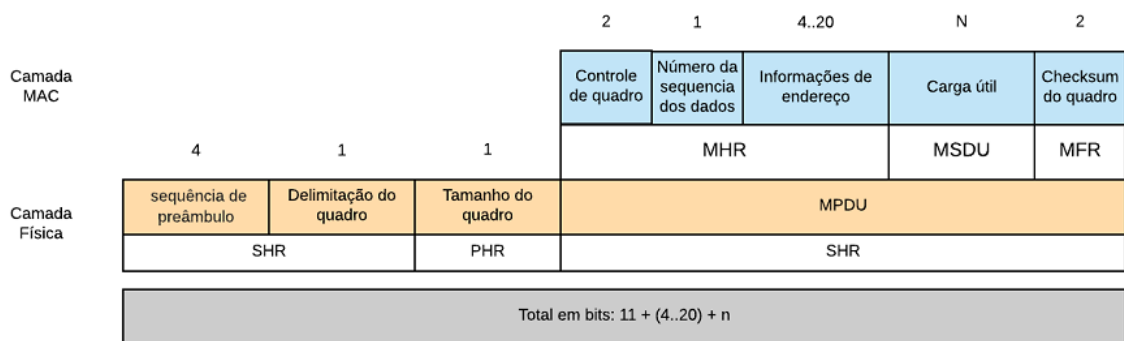


Figura 8 – Quadro 802.15.4

A taxa de transmissão é considerada como $0.26 \mu\text{bits/s}$, a corrente elétrica para recepção e escutar o meio é considerada 7.0mA enquanto a corrente elétrica para a transmissão é fixada em 21.5mA .

Será considerado o formato de quadro de camada física 802.15.4 (Figura 8) Como proposto em [ADAMS 2006]. como observado nos modelos acima, a energia gasta para

transmissão ou recepção dependem do tamanho dos dados. Nos experimentos realizados com PCASW foi considerado o tipo de dados Float, que compreende valores entre 10^{-38} e 10^{38} , utiliza 4 bytes, e tem a precisão de 7 dígitos. São consideradas as variáveis de temperatura, humidade, luminosidade e energia.

5 Resultados

Esta seção descreve alguns detalhes da execução do algoritmo bem como os resultados atingidos em termos de energia gasta e erro associado ao processo, a fim de investigar redução de dados com PCASW.

Foram executadas simulações com 54 datasets de nós diferentes, cada um com cerca de 40mil medições. Alguns parâmetros utilizados nas simulações são apresentados na tabela 2.

Tabela 2 – Parâmetros da Simulação

Parâmetros	descrição	Valor
n	Número de dimensões participantes do processo	4
k	Número de dimensões Resultantes do processo	1
t	Tamanho da Janela Mínimo	10
tmax	Tamanho máximo da Janela	300
var(k)	Variância total mínima em k	90%

5.1 Erro

Os resultados da simulação podem ser vistos na Figura 9, é possível observar que a taxa de erro média para cada medição é de aproximadamente 0.1774493, além disso o erro se manteve controlado, nunca sendo superior a 0.4 por medição.

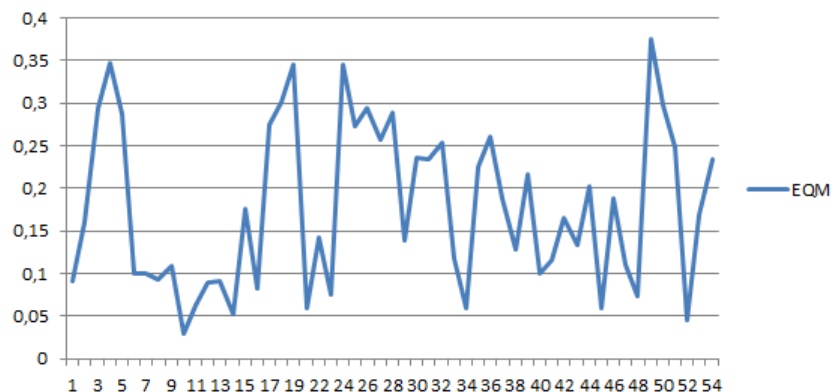


Figura 9 – Erro médio por medição em cada Nó

Como visto, os níveis de erro se mantiveram aceitáveis em todas as simulações considerando que houve uma redução de aproximadamente 75% dos dados originais.

5.2 Energia

Em termos de energia, os resultados podem ser verificados no gráfico da figura 10, em vermelho a energia gasta sem a utilização do PCASW e em azul a energia gasta com a transmissão dos dados utilizando nossa abordagem. É visível a diferença entre os valores, como descrito nos capítulos anteriores, PCASW proporciona uma diminuição de até $1/n$, sendo n o número de dimensões participantes do processo, como nas simulações foram utilizadas 4 dimensões, a redução média para cada sensor, ficou em torno de 75%.

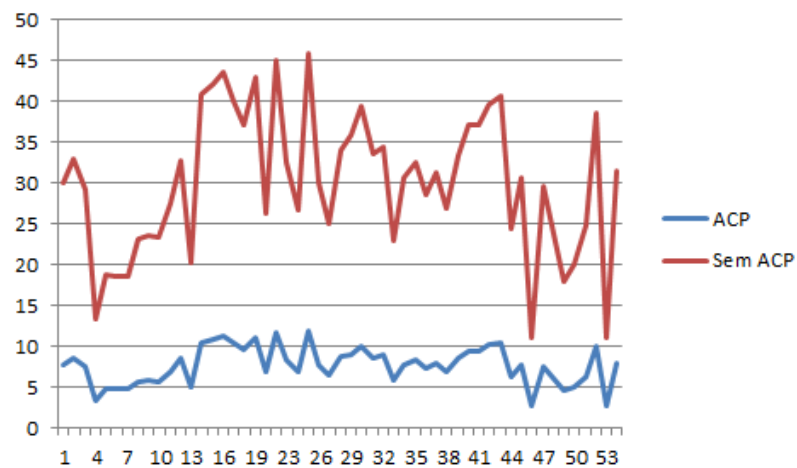


Figura 10 – Energia gasta com e sem PCASW

6 Conclusão e Trabalhos Futuros

6.1 Conclusão

Após os resultados encontrados foi possível verificar a efetividade do algoritmo proposto em relação aos objetivos impostos anteriormente. Em termos de energia gasta pelos nós sensores a partir da transmissão de informação, os objetivos foram alcançados obtendo uma redução de 4 para somente uma dimensão nos dados utilizados em testes, diminuindo a informação que trafega pela rede, e por consequência, aumentando a vida útil de toda a rede. Além disto o erro se manteve controlado e relativamente baixo, preservando assim, as características analíticas dos dados enviados.

6.2 Trabalhos Futuros

Como trabalhos futuros, iremos verificar a eficácia da abordagem proposta em outros cenários, em ambientes de simulação de redes inteiras a fim de verificar os impactos na vida útil da rede.

Testar a utilização em conjunto com outras técnicas, como protocolos de roteamento cientes de energia, a fim de diminuir a sobrecarga dos nós intermediários [BRITTES 2007], problema recorrente neste tipo de rede.

Referências

- ADAMS, J. T. An introduction to ieeb std 802.15.4. IEEE Aerospace Conference, 2006. Citado na página 24.
- AKYILDIZ, I. F. et al. A survey on sensor networks. *IEEE communications magazine*, IEEE, v. 40, n. 8, p. 102–114, 2002. Citado na página 15.
- AQUINO, A. E. L. L. D. *Redução de dados em redes de sensores sem fio baseada em stream de dados*. Tese (Doutorado) — Tese de doutorado, Universidade Federal de Minas Gerais, 2008. Citado na página 12.
- BEZERRA, V. et al. A quality-aware and energy-efficient context management framework for ubiquitous systems. In: IEEE. *Advanced Information Networking and Applications (AINA), 2014 IEEE 28th International Conference on*. [S.l.], 2014. p. 568–575. Citado 3 vezes nas páginas 17, 21 e 24.
- BOGENA, H. et al. Potential of wireless sensor networks for measuring soil water content variability. *Vadose Zone Journal*, UNKNOWN, v. 9, n. 4, p. 1002–1013, 2010. Citado na página 15.
- BRITTES, M. *Uma proposta para melhoria de desempenho do protocolo leach para rssf*. Tese (Doutorado) — Dissertação de mestrado, Universidade tecnológica federal do Paraná, Curitiba, 2007. Citado na página 28.
- CAMILO, T. et al. An energy-efficient ant-based routing algorithm for wireless sensor networks. In: SPRINGER. *International Workshop on Ant Colony Optimization and Swarm Intelligence*. [S.l.], 2006. p. 49–59. Citado na página 12.
- CARVALHO, C. G. N. D. et al. Multiple linear regression to improve prediction accuracy in wsn data reduction. In: IEEE. *Network Operations and Management Symposium (LANOMS), 2011 7th Latin American*. [S.l.], 2011. p. 1–8. Citado na página 18.
- DIAS, G. M.; BELLALTA, B.; OECHSNER, S. A survey about prediction-based data reduction in wireless sensor networks. *ACM Computing Surveys (CSUR)*, ACM, v. 49, n. 3, p. 58, 2016. Citado na página 12.
- HENNING, M.; FONSECA, M.; MUNARETTO, A. Pbp-um protocolo baseado em políticas para controle eficiente de energia em redes de sensores sem fio. Citado na página 16.
- JACKSON, J. E. A user's guide to principal components. ISBN 978-0-471-47134-9., 2003. Citado na página 20.
- LEMOES M., d. C. C. L. D. R. R.; FILHO, R. H. Reducing energy consumption in provisioning of virtual sensors by similarity of heterogeneous sensors. IEEE 31st International Conference on Advanced Information Networking and Applications (AINA), 2017. Citado na página 18.
- LOUREIRO, A. A. et al. Redes de sensores sem fio. In: SN. *Simpósio Brasileiro de Redes de Computadores (SBRC)*. [S.l.], 2003. p. 179–226. Citado na página 15.

MARGI, D. A. G. O. e C. B. Roteamento ciente de energia em redes de sensores sem fio definidas por software. XXXV SIMPOSIO BRASILEIRO DE TELECOMUNICAÇÕES E PROCESSAMENTO DE SINAIS - SBrT2017., 2017. Citado na página [12](#).

TEIXEIRA, I. Roteamento com balanceamento de consumo de energia para redes de sensores sem fio. *Rio de Janeiro*, 2005. Citado na página [17](#).

WIKIPEDIA. *R (linguagem de programação)*. Citado na página [23](#).