

Continuous Monitoring in WSN: A Fuzzy-Probabilistic Approach

Flávio R. S. Nunes, José E. B. Maia

¹CCT - Universidade Estadual do Ceará (UECE)
60714-903 - Fortaleza - Ceará - Brasil

flavio.r.sn@gmail.com, jose.maia@uece.br

Abstract. *This work presents and evaluates a fuzzy-probabilistic strategy to save energy in Wireless Sensor Networks (WSNs). The energy savings are obtained with the sensor nodes, no longer sensing and transmitting measurements. In this simple strategy, in each epoch each sensor node transmits its measurement with probability p , and does not transmit with probability $(1 - p)$, does not correlate with that of any other sensor node. The task at the sink node, which is to estimate the sensor field at non-sensed points, is solved using fuzzy inference to impute the non-transmitted data followed by regression or interpolation of the sensed scalar field. In this, Nadaraya-Watson regression, regression with Fuzzy Inference and Radial Base Functions Interpolation are compared. The compromise curve between the value of p and the accuracy of the sensor field estimation measured by root mean square error (RMSE) is investigated. When compared to a published linear prediction strategy of the literature, the results show a small loss of performance versus the great simplification of the procedure in the sensor node, making it advantageous in applications that require extremely simple network nodes.*

1. Introdução

Uma rede de sensores sem fio (RSSF) é caracterizada por uma quantidade de nós-sensores com capacidades para sensoriar o ambiente, processar e armazenar dados, e comunicar-se com outros nós-sensores [Akyildiz et al. 2002]. O que distingue uma RSSF de outras redes de computadores sem fio em geral é, principalmente, que essas capacidades são reduzidas nas RSSF. Entretanto, a limitação mais restritiva em algumas aplicações das RSSF é a energia disponível nos nós sensores, que por sua vez define o tempo de vida da rede. Considera-se o caso onde os nós não podem ser alimentados por energia contínua e a substituição de baterias é inviável [Rault et al. 2014].

Para economizar energia, os nós sensores estão capacitados a alternarem entre os estados ativo ou hibernando [Rawat et al. 2014]. No estado de hibernação, a atividade do nó é reduzida a um mínimo suficiente para decidir quando voltar ao estado ativo. Por exemplo, quando esta operação é cíclica apenas um *timer* pode ficar em atividade, ainda assim de baixo consumo. Quando ativo, a energia em um nó sensor é consumida nas atividades de sensoriamento, processamento, recepção e transmissão de mensagens.

Embora a tecnologia venha permitindo o contínuo aumento da capacidade de processamento dos nós sensores, considera-se neste trabalho a classe de aplicações de monitoramento contínuo [Yick et al. 2008] por uma RSSF densa composta de centenas de nós com baixas capacidades de processamento, armazenamento e energia. Esse cenário

exige algoritmos simples e ciclos de baixa atividade dos nós sensores. A exigência de nós sensores de muito baixa capacidade pode ser originada pela necessidade de nós de dimensões muito reduzidas. A discrição ou disfarce da aplicação, como em aplicações militares, ou um reduzido espaço para acomodar os nós pode originar essa necessidade. [Đurišić et al. 2012, Narayanan et al. 2016, Bdiri and Derbel 2015, Xiao et al. 2017].

O monitoramento contínuo por RSSF densa implica na possibilidade de redundâncias espacial e temporal nos dados sensoriados [Vuran et al. 2004]. Essas propriedades dos dados podem ser aproveitadas para reduzir o ciclo ativo dos nós como também suas atividades durante o estado ativo. Para isso, frequentemente são utilizadas técnicas de regressão [de Carvalho et al. 2011, AlShawi et al. 2012]. Note entretanto que os parâmetros de um tal algoritmo devem ser cuidadosamente selecionados pois a subamostragem espacial ou temporal do campo sensor implica em um compromisso entre a acurácia da reconstrução do campo sensor e a duração de vida da RSSF via o consumo de energia pelos nós [Bouabdallah et al. 2009]. Diversos algoritmos sofisticados já foram propostos e avaliados na literatura para prolongar a vida útil de RSSF utilizando este compromisso [Hodge et al. 2015].

Visando nós sensores de baixa capacidade, este trabalho propõe e avalia um algoritmo simples, de baixo processamento nos nós, para prolongar a vida útil de RSSF densas baseado no compromisso entre acurácia e taxa de amostragem. A reconstrução do campo sensoriado é baseada em inferência fuzzy. O desempenho, utilizando a curva erro de reconstrução do campo sensor versus número de mensagens transmitida, é comparado com outro algoritmo descrito na literatura, baseado em predição linear [Maia et al. 2013].

O restante deste artigo está assim organizado. A Seção 2 analisa alguns trabalhos relacionados relevantes e a Seção 3 revisa os conceitos necessários ao entendimento do trabalho. A Seção 4 mostra o procedimento de construção do modelo de inferência fuzzy proposto, baseado em conhecimento do domínio. A Seção 5 é de resultados, seguida pela Seção 6 de conclusão.

2. Trabalhos Relacionados

Há ampla literatura propondo técnicas para reduzir o consumo de energia das RSSF em aplicações de monitoramento contínuo. Elas incluem agregação de dados, agrupamento de nós para transmissão, sensoriamento compressivo, redução de transmissões baseada em modelos preditivos e escalonamento de nós. Essa seção faz uma breve revisão de trabalhos diretamente relacionados a esta pesquisa. Revisões amplas do assunto podem ser encontrada em [Rault et al. 2014, Ghate and Vijayakumar 2018, Khan et al. 2015].

As técnicas citadas não são mutuamente excludentes. O foco desta pesquisa é na redução do número de mensagens geradas. Os resultados serão comparados com uma técnica que utiliza um modelo preditivo para também evitar a geração e transmissão de dados sensoriados. Um apanhado destas técnicas pode ser encontrado em [Maia et al. 2013].

Em [Diwakaran et al. 2019], usando o fato de que sensores adjacentes mostram alta correlação espacial e temporal, é desenvolvido um modelo preditivo para conservação de energia centrado nos dados já coletados, para reduzir a transmissão de dados redundantes. Agrupamentos são formados e modelos de regressão são contruídos para representar o agrupamento. Medições que podem ser previstas dentro de certo erro máximo não são

transmitidas. Uma curva de desempenho erro quadrático versus economia de energia é apresentada.

Deligiannakis et al. [Deligiannakis and Kotidis 2006] desenvolvem um algoritmo o qual denominam Self-Based Regression (SBR) para ser aplicado em processamento de consultas eficiente em energia. Modelos de regressão são construídos *in network* para os dados em grupos de sensores e o desempenho é medido pela redução da quantidade de dados transmitidos entre os nós vizinhos.

Em Zhang et al. [Zhang et al. 2013] os sensores são divididos em diferentes agrupamentos onde, em cada grupo somente um nó de referência funciona amostrando periodicamente a grandeza sensoriada enquanto que os demais nós no agrupamento realizam sensoriamento compressivo (CS). O nó sink utiliza um algoritmo de predição para estimar as séries temporais de todos os nós baseado na séries de sinais recebidas dos nós de referência. O desempenho é o erro de reconstrução.

A ideia central nos trabalhos [Matos et al. 2010, Maia et al. 2013] é que o nó sink constroi modelos preditivos lineares para a série temporal em cada nó sensor e envia os coeficientes do modelo para cada nó simular a predição realizada pelo sink antes de transmitir. Se a predição cai dentro de um limiar de erro estabelecido, o nó não transmite, economizando energia. Sempre que o erro é maior que o limiar considera-se que o modelo desatualizou-se e o sink recalcula e re-envia os coeficientes para aquele nó.

Duas características dos trabalhos citados são que eles requerem o cálculo e atualização de modelos preditivos, seja no sink seja nos nós sensores, e que coeficientes de modelos precisam de serem enviados pelo nó sink ou trocados entre nós de um agrupamento gerando mensagens de controle. Esses pontos motivaram o presente trabalho no qual uma estratégia simples, totalmente distribuída e que não requer mensagens de controle é proposta. O desempenho é comparado com o trabalho [Maia et al. 2013]. Este trabalho foi escolhido para comparação pela similaridade de configuração do problema e por apresentar informações para sua completa reprodutibilidade.

3. Algoritmos

Nos modelos de monitoramento analisados a seguir existe dois componentes: o algoritmo de decisão de sensoriamento e transmissão pelo nó sensor, e o algoritmo de reconstrução do campo sensor pelo nó sink. Serão analisados dois algoritmos de decisão e três algoritmos de reconstrução. Eles podem ser combinados gerando seis configurações possíveis.

3.1. Estratégias proposta e alternativa

A estratégia proposta neste trabalho será chamada Modelo Probabilístico. O seu desempenho será comparado com outra chamada Modelo Predição, apresentado em [Matos et al. 2010, Maia et al. 2013]. Uma breve revisão de predição baseada em regressão linear é apresentada na Seção 2.2 para tornar esse texto autocontido. Embora não sincronizados, os nós sensores são programados para ciclicamente medir e transmitir as medições. Uma época é uma rodada de medições. Os comportamentos dos nós sensor e *sink* em cada uma delas é como descrito a seguir.

Comportamento de um nó sensor: no Modelo Probabilístico, o nó simplesmente sensoria (mede) e transmite em cada época com probabilidade p . No Modelo Predição,

o nó sensor recebe do nó sink os coeficientes de um modelo para fazer a predição da próxima medição. Em cada época o nó utiliza o último modelo de regressão linear recebido para fazer a predição do dado. Se a diferença entre o valor sensoriado e o valor predito for maior que um certo *threshold*, ele transmite, caso contrário, não.

Comportamento do nó *sink*: no Modelo Probabilístico o nó *sink* sempre completa os dados de uma época adotando inferência fuzzy para imputar os dados não transmitidos de cada nó. Para avaliação da acurácia dessa estratégia, uma superfície de regressão não linear por inferência fuzzy ou Nadaraya-Watson, ou de interpolação com funções base radial é construída com os dados reais e outra com os dados adotados. O erro entre essas duas superfícies nos pontos dos sensores é a medida de acurácia.

No Modelo Predição, o nó *sink* compõe uma época com os dados transmitidos, para os nós que transmitiram naquela época, e calcula e adota os dados preditos, para os nós que não transmitiram naquela época. Sempre que o nó *sink* recebe dados de um nó sensor, ele recalcula os coeficientes da regressão linear para aquele nó e os envia ao respectivo nó. Isso é feito porque, se um nó sensor transmitiu é porque seu modelo desatualizou-se em prever as medições. A avaliação da acurácia é feita da mesma forma que para o Modelo Probabilístico.

3.2. Predição linear

A Regressão Polinomial univariada modela a relação entre duas variáveis ajustando uma equação aos dados observados. A Equação (1) apresenta a função de ajuste univariada:

$$y = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \dots + \alpha_m x^m + \epsilon, \quad (1)$$

onde $\alpha_0 \dots \alpha_m$ são coeficientes polinomiais, $x \dots x^m$ são potências da variável independente, ϵ é uma variável que inclui todos os fatores residuais mais os possíveis erros de medição.

Para este trabalho utilizou-se regressão linear simples como em [Maia et al. 2013]. A solução de redução de dados na RSSF aplica-se fazendo uma regressão simples, com mínimos quadrados, utilizando como variável independente para a predição o tempo ou época multiplicada pelo coeficiente α_1 somando-se com intercepto α_0 . Como resultado obtém-se a estimacão \hat{y} , e deste modo consegue-se determinar o erro a partir do valor predito e do valor real. Isso está mostrado na Equação (2):

$$\hat{y} = \alpha_0 + \alpha_1 x. \quad (2)$$

Desde que se tenha o valor do erro, ele é então comparado a um limiar, e cada nó de posse deste valor compara e decide o envio ou não do conjunto de dados para o nó *sink*.

3.3. Regressão Nadaraya-Watson (NW)

Um estimador de regressão de Nadaraya-Watson (ERNW) é baseado em regressão de núcleo (*kernel*) fundamentada em estimacão de densidade [Haykin 2000]. Ela pode ser definida utilizando uma função de ponderacão normalizada $W_{N,i}$ dada por

$$W_{N,i}(x) = \frac{K\left(\frac{x-x_i}{h}\right)}{\sum_{j=1}^N K\left(\frac{x-x_j}{h}\right)}, i = 1, 2, \dots, N \quad (3)$$

com

$$\sum_{i=1}^N W_{N,i}(x) = 1, \text{ para todo } x, \quad (4)$$

em que $K(\cdot)$, na Equação (3), é um núcleo com propriedades que se assemelham com as de uma função de densidade de probabilidade e o termo h , chamado de largura de banda, define a abertura do núcleo. De forma resumida escreve-se a função de estimação de regressão de núcleo F na Equação (5) que é uma média ponderada dos valores observados y_i :

$$y = F(x) = \sum_{i=1}^N W_{N,i}(x)y_i \quad (5)$$

Assim obtém-se o valor estimado y para um ponto medido x .

3.4. Interpolação com Função Base Radial (RBF)

Interpolação com funções de base radial (RBF) é um método primário para reconstrução de uma função multivariada desconhecida a partir de alguns pontos de dados [Haykin 2000]. Existem várias funções de base radial que podem ser adotadas. Neste trabalho foi utilizada a mais frequente que é a RBF Gaussiana.

No método RBF, dado um conjunto de n pontos de dados distintos x_j da variável independente e os correspondentes valores medidos y_j , a função de interpolação RBF é dada por

$$f(x) = \sum_j \lambda_j \phi(x - x_j), \quad (6)$$

onde $\phi(r)$, $r > 0$, é uma função base radial, sendo nesse trabalho a função Gaussiana dada por

$$\phi(r) = \exp(-\epsilon r^2). \quad (7)$$

Os coeficientes da expansão λ_j são calculados aplicando as condições de interpolação $f(x_j) = y_j$ as quais levam a seguinte sistema linear simétrico de ordem n

$$\mathbf{A}\lambda = \mathbf{y}, \quad (8)$$

onde as entradas a_{ij} da matriz \mathbf{A} são dadas por $a_{ij} = \phi(x_i - x_j)$. As condições para que a matriz \mathbf{A} seja não singular [Micchelli 1984] são atendidas pela RBF Gaussiana.

4. Regressão com Inferência fuzzy (IF)

Note que inferência fuzzy é utilizada duas vezes: uma vez para imputar valores perdidos ou não transmitidos pelos nós sensores, ou para substituição de outliers detectados, e outra vez para a estimação do campo sensor em pontos de consultas. A concepção do SIF é a mesma nos dois casos, assim será apresentada apenas uma descrição e apontadas as diferenças entre os dois casos, ao final.

Para inferência fuzzy foi utilizado um modelo de mandami [Jang et al. 1997]. Um sistema fuzzy baseado no modelo de Mandami pode ser dividido em 4 partes: fuzzificação das variáveis de entrada, regras de avaliação, agregação das regras de saída e defuzzificação.

O modelo construído foi baseado em conhecimento do especialista do domínio. A abordagem proposta foi avaliada utilizando o *Intel Lab data set* de domínio público contendo medições de 54 sensores realizadas no Intel Berkeley Research Lab. Para explicar a construção do sistema de Mandami, a Figura 1 mostra um mapa da distribuição dos sensores. É com base nele que o sistema de inferência será construído.

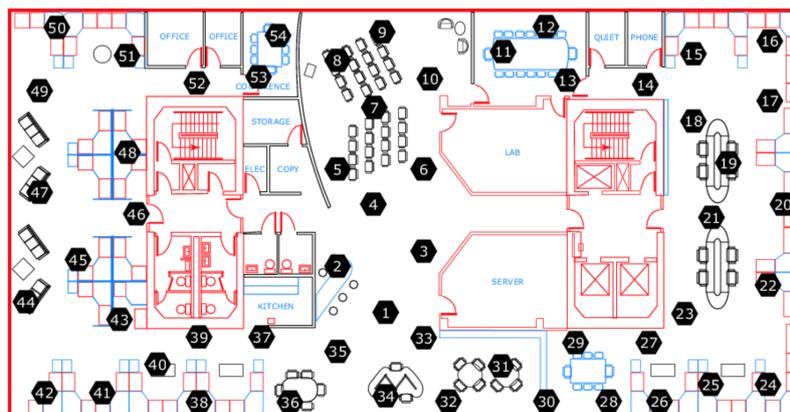


Figura 1. Distribuição dos sensores no laboratório

4.1. Fuzzificação das variáveis de entrada

Uma variável linguística 'temperatura' foi criada para cada nó sensor. Analisando-se testes experimentais adotaram-se três termos: 'frio', 'médio' e 'quente', com funções de pertinência sigmoide decrescente, Gaussiana e sigmoide crescente, respectivamente. Uma vantagem dessas funções escolhidas é a dependência de apenas dois parâmetros em cada uma.

Para se obter os parâmetros aplicáveis às funções de pertinência analisou-se o gráfico de todas as medições de temperatura contidas no conjunto de dados. Um exemplo de dados, para o sensor 10, está mostrado na Figura 2.

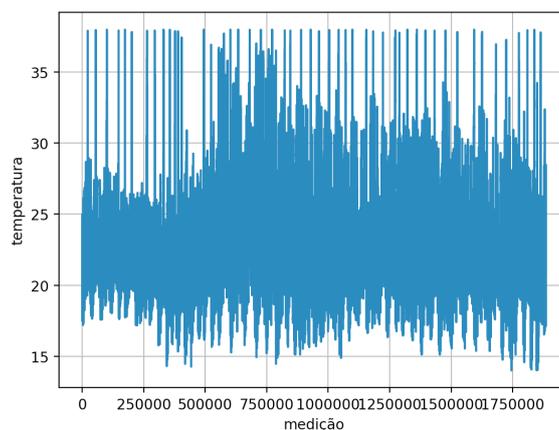


Figura 2. Análise das constantes das funções de pertinência

Para simplificação, foram adotadas as mesmas funções de pertinência com os mesmos parâmetros para todos os sensores e também para antecedentes e consequentes. Isso

reduziu substancialmente o número de parâmetros a sintonizar, que foram obtidos por *cross validation*. Com as constantes escolhidas, as funções de pertinência geradas podem ser vistas na Figura 3. Os parâmetros obtidos foram 19,86 e -4,84 para 'frio', 21,68 e 2,95 para 'médio', e 37,24 e 8,33 para 'quente'.



Figura 3. Funções de pertinência “quente”, “médio” e “frio”

4.2. Regras de Inferência, Agregação e Defuzzificação

```
dependencies = [
    [54, [53]],
    [53, [54]],
    [52, [51]],
    [51, [50, 52]],
    [50, [49, 51]],
    [49, [50, 51]],
    [48, [46, 47]],
    [47, [49, 45]],
```

Figura 4. Algumas das dependências dos sensores para gerar as regras

Um sistema de inferência fuzzy é criado para estimar a temperatura em cada nó sensor. As regras de inferência fuzzy foram construídas *ad hoc* examinando o domínio da Figura 1. Sendo que o número de regras cresce exponencialmente com o número de variáveis independentes, foi imposta a restrição de que o consequente de cada nó sensor depende apenas de dois outros nós sensores, tomados como antecedentes. Assim, cada conjunto de regras possui no mínimo 3 e no máximo 9 regras. Alguns testes permitindo um número maior de dependências não resultaram em ganho significativo. A Figura 4 mostra exemplos de dependências obtidas.

Para um exemplo do procedimento *ad hoc* utilizado, note, pela Figura 1, que os sensores 54 e 53 estão em uma mesma sala do ambiente construído. Com base nisso, as dependências dos sensores 54 e 53 foram criadas: 54 depende de 53 e 53 depende de 54, como mostrado nas duas primeiras linhas da Figura 4. O sensor 50, por outro lado, encontra-se em um compartimento bastante aberto, o que possibilita que dependa de um maior número de sensores. Foram escolhidos os dois sensores mais próximos: 49 e 51. Os mesmos padrões exemplificados foram utilizados para criar o restante das dependências não citadas, resultando assim em 52 conjunto de regras.

O operador de agregação utilizado foi a composição max-min e para a etapa de defuzzificação foi utilizado o método do centróide. Outros operadores de agregação e defuzzificação foram testados mas não geraram melhores resultados.

Por fim, a Figura 5 mostra a superfície de reconstrução gerada pelo algoritmo Inferência Fuzzy (IF) para uma época específica no Modelo Probabilístico, a partir da qual o valor da temperatura pode ser estimada em um ponto qualquer (x,y) fora das localizações dos nós sensores.

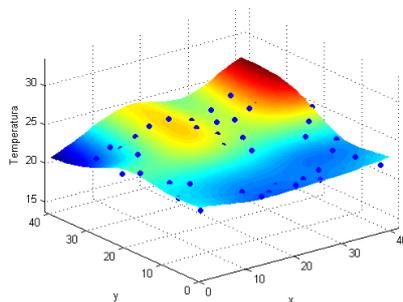


Figura 5. Campo escalar reconstruído para a época 500 utilizando Inferência Fuzzy.

Finalmente, o mesmo SIF descrito é utilizado também na reconstrução do campo para estimar valores em pontos de consulta. A diferença é que na reconstrução do campo, dado um ponto de consulta, o sistema identifica os dois sensores mais próximos daquele ponto usando distância euclidiana, e utiliza seus valores no SIF para inferir o terceiro.

5. Resultados e discussão

5.1. Dados e Pré-processamento

A abordagem proposta foi avaliada utilizando o *Intel Lab data set*¹ de domínio público contendo medições de 54 sensores realizadas no Intel Berkeley Research Lab. Os nós sensores coletaram medições de umidade, temperatura, luminosidade e energia a cada 31 segundos. Apenas dados de temperatura foram utilizados nestes experimentos. Sendo que dois dos sensores não funcionaram corretamente (os de *id* 5 e 15), foram utilizados dados de 52 sensores. Os pontos de dados faltantes foram identificados pela informação de timestamp sabendo-se que cada sensor gera uma medição a cada 31 s.

Este *dataset* possui uma quantidade *outliers* e de dados perdidos requerendo uma fase de pré-processamento. Foram utilizados métodos simples para detectar *outliers* e imputar os dados faltantes uma vez que isso não está em avaliação e não tem consequência direta no procedimento comparativo. O procedimento consistiu em suprimir os *outliers* detectados e tratá-los como dados faltantes. Os dados faltantes foram imputados por simples repetição do último valor válido e os *outliers* foram detectados por uma heurística que consistiu em considerar *outlier* toda medição com variação maior que 10% do valor da medição anterior. Essa heurística é justificada pela superamostragem da temperatura ambiental.

Após o pré-processamento, um total de 20 conjuntos de dados (*datasets*) foram construídos para os experimentos, cada um com dados relativos a cinco dias de medição. Cada conjunto foi definido escolhendo aleatoriamente um ponto de início no *dataset* original e tomando os dados de cinco dias a partir daquele ponto. Os dados relativos a cinco dias de coleta foram divididos em dois segmentos, um segmento de projeto, correspondente a três dias, e um segmento de teste, para avaliação de desempenho, correspondente a dois dias. Em cada conjunto de projeto foi utilizada *cross validation* para selecionar os parâmetros do modelo. O modelo com melhor resultado foi escolhido. Os resultados

¹Intel Berkeley Laboratory Data, <http://berkeley.intel-research.net/labdata/>

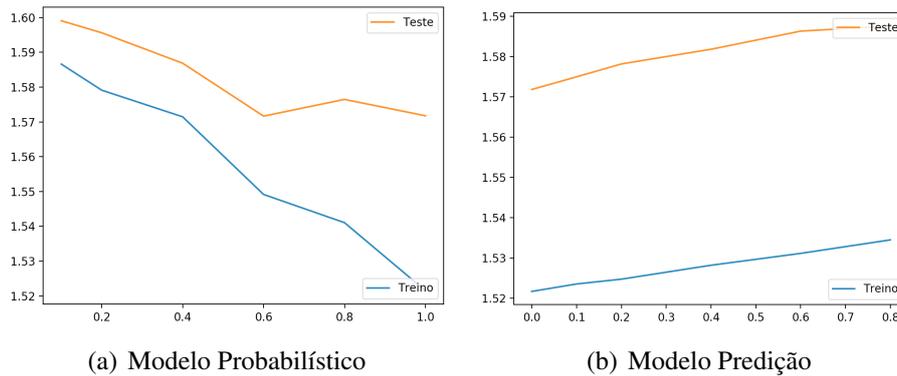


Figura 6. Erro de reconstrução (RMSE) utilizando Inferência Fuzzy.

mostrados na próxima seção referem-se a aplicação deste modelo aos vinte conjuntos de teste.

5.2. Resultados

Os resultados desta seção estão organizados em uma tabela e alguns gráficos. Foram comparados entre si os desempenhos das seis combinações entre mecanismo de decisão e algoritmo de regressão (reconstrução). Os resultados para o método proposto são mostrados na Tabela 1.

É de se notar o seguinte na Tabela 1. Que os valores da última coluna se repetem para os modelos Predição e Probabilístico uma vez que $th = 0$ ou $r = 1$ são equivalentes a $p = 1.00$. Que o RMSE para a reconstrução RBF é superior aos outros métodos no dois modelos e assim não é a melhor opção. Que embora o RMSE para NW seja ligeiramente menor que para IF nos dois modelos de transmissão, o seu desvio padrão é de três a quatro vezes maior significando menor previsibilidade do resultado. Assim sendo pode-se concluir desta tabela que a combinação Modelo Probabilístico de sensoriamento, transmissão e inferência fuzzy para imputação de dados e para reconstrução do campo escalar fornece a melhor combinação com baixo processamento, baixo consumo de energia por sensoriamento e transmissão de mensagens e maior previsibilidade da acurácia dos resultados da estimação do campo sensoriado.

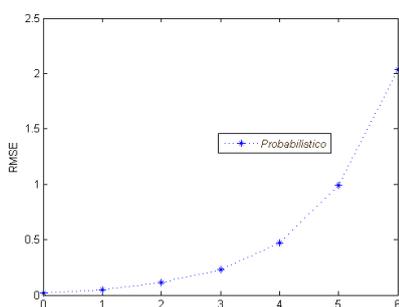
Esta tabela mostra claramente que não há perda do modelo probabilístico em relação ao modelo de predição apresentado em [Maia et al. 2013] para os dados testados. Entretanto, no modelo probabilístico o nó sensor não realiza medição e nem calcula a predição a cada época para decidir como também não exige mensagens de transmissão dos coeficientes dos modelos de regressão de cada nó sensor. Para comparação, simulações foram realizadas utilizando regressão Nadaraya-Watson (NW) e interpolação Base Radial (RBF) no lugar da Inferência Fuzzy.

Para controle de *overfitting* dos experimentos a Figura 6 mostra o RMSE médio dos conjuntos de treinamento e de teste para os dois modelos com IF. Na figura 6(a) o parâmetro no eixo das abcissas é p para o modelo probabilístico e na 6(b) é th para o modelo predição.

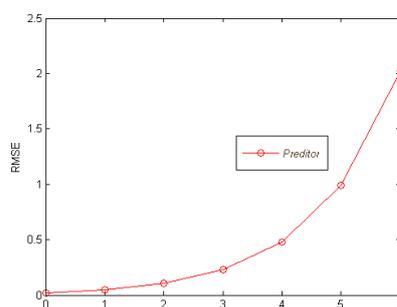
Considerou-se também a hipótese de que a temperatura ambiental está superamostrada na taxa de uma amostra a cada 31 s. Para tanto foram realizados experimentos para

Tabela 1. Média (m) e desvio padrão (dp) do RMSE da comparação entre os modelos de transmissão probabilístico e com predição para os três modelos de reconstrução (IF, NW e RBF).

Mod.Prob.	p	-	0.10	0.20	0.40	0.60	0.80	1.00
	NW	m	1,599	1,595	1,586	1,571	1,576	0,906
		dp	0,0435	0,0486	0,0366	0,0349	0,0270	0,0243
	IF	m	1,738	1,586	1,385	1,291	1,186	0,869
		dp	0,095	0,095	0,091	0,097	0,090	0,097
	RBF	m	1,777	1,654	1,476	1,408	1,368	1,204
		dp	0,089	0,088	0,078	0,081	0,092	0,095
Mod.Pred.	th	-	2.00	1.00	0.40	0.20	0.10	0.00
	r	-	0,22	0,49	0,82	0,88	0,97	1
	NW	m	1,587	1,586	1,581	1,578	1,575	0,906
		dp	0,0235	0,0221	0,0276	0,0262	0,0463	0,0243
	IF	m	1,478	1,336	1,285	1,201	1,000	0,869
		dp	0,084	0,088	0,092	0,092	0,095	0,097
	RBF	m	1,528	1,421	1,455	1,346	1,321	1,204
		dp	0,089	0,076	0,082	0,099	0,085	0,095



(a) Modelo Probabilístico



(b) Modelo Predição

Figura 7. Erro de reconstrução (RMSE) com os dados subamostrados (IF).

verificar a sensibilidade dos resultados à frequência de amostragem da temperatura. Isso foi simulado suprimindo na forma intercalada uma fração cada vez maior das amostras no *data set*. Dessa forma tomou-se uma a cada 2^n amostras para $n = \{1, 2, 3, 4, 5, 6\}$. Os resultados são mostrados na Figura 7(a) para o modelo probabilístico e na Figura 7(b) para o modelo de predição. Note que, em ambos os casos, a partir de $n = 3$, ou seja, uma amostra a cada $8 \times 31 = 248 = 4, 14$ minutos, o RMSE cresce rapidamente sendo este um valor limite. Nota-se assim que a regressão espacial entre sensores está beneficiando-se da redundância temporal dos dados.

6. Conclusão

O trabalho avaliou as duas componentes da estratégia para prolongar a vida das RS-SFs: o mecanismo de decisão probabilístico descentralizado com inferência fuzzy para imputação de dados e o algoritmo de reconstrução baseado em inferência nebulosa (*fuzzy*).

O desempenho do mecanismo de decisão probabilístico foi comparado com aquele de uma estratégia de predição na rede (*prediction in-network*) publicada na literatura e a regressão por inferência nebulosa foi comparada com o regressor de Nadaraya-Watson (NW) e interpolação RBF.

Os desempenhos das seis combinações de mecanismo de decisão com mecanismo de regressão foram analisados utilizando um conjunto de dados publicamente disponível. O desempenho do conjunto foi medido pelo RMSE de reconstrução do campo escalar. Os resultados mostram evidência suficiente para que se possa recomendar a combinação de decisão probabilística com inferência nebulosa. A vantagem da inferência nebulosa pode ser atribuída ao uso de conhecimento do domínio na construção do modelo e a do mecanismo probabilístico de decisão é ele não requerer qualquer tipo de sobrecarga em mensagens de controle e nem de processamento no nó sensor. A decisão probabilística é extremamente simples, podendo ser tomada com o nó sensor no estado de hibernação, sendo apropriada para os nós sensores de muito baixa capacidade de processamento e de dimensões muito reduzidas exigidos por algumas aplicações.

O trabalho deixa algumas linhas de pesquisa abertas para o futuro. O foco deste trabalho foi o número de mensagens originais geradas pela rede. É necessário investigar como esse mecanismo interage com os demais mecanismos do funcionamento de uma RSSF tais como o roteamento de mensagens, agrupamento de nós sensores (*clustering*) e movimentação do nó sorvedouro (*sink*). É necessário também ampliar os testes para novos conjuntos de dados e validar em uma RSSF experimental uma vez que os resultados foram obtidos por simulação.

Referências

- Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., and Cayirci, E. (2002). A survey on sensor networks. *IEEE communications magazine*, 40(8):102–114.
- AlShawi, I. S., Yan, L., Pan, W., and Luo, B. (2012). Lifetime enhancement in wireless sensor networks using fuzzy approach and a-star algorithm. *IEEE Sensors Journal*, 12(10):3010–3018.
- Bdiri, S. and Derbel, F. (2015). An ultra-low power wake-up receiver for realtime constrained wireless sensor networks. In *Proceedings of the AMA Conferences, Nuremberg, Germany*, pages 19–21.
- Bouabdallah, N., Rivero-Angeles, M. E., and Sericola, B. (2009). Continuous monitoring using event-driven reporting for cluster-based wireless sensor networks. *IEEE Transactions on Vehicular Technology*, 58(7):3460–3479.
- de Carvalho, C. G. N., Gomes, D. G., de Souza, J. N., and Agoulmine, N. (2011). Multiple linear regression to improve prediction accuracy in wsn data reduction. In *2011 7th Latin American Network Operations and Management Symposium*, pages 1–8. IEEE.
- Deligiannakis, A. and Kotidis, Y. (2006). Exploiting spatio-temporal correlations for data processing in sensor networks. In *International conference on GeoSensor Networks*, pages 45–65. Springer.
- Diwakaran, S., Perumal, B., and Devi, K. V. (2019). A cluster prediction model-based data collection for energy efficient wireless sensor network. *The Journal of Supercomputing*, 75(6):3302–3316.

- Ghate, V. V. and Vijayakumar, V. (2018). Machine learning for data aggregation in wsn: A survey. *International Journal of Pure and Applied Mathematics*, 118(24):1–12.
- Haykin, S. (2000). *Redes neurais: princípios e prática*. Bookman, 2th edition.
- Hodge, V. J., O’Keefe, S., Weeks, M., and Moulds, A. (2015). Wireless sensor networks for condition monitoring in the railway industry: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 3(16):1088–1106.
- Jang, J. S. R., Sun, C. T., and Mizutani, E. (1997). *Neuro-Fuzzy and Soft Computing—A Computational Approach to Learning and Machine Intelligenc*. Prentice-Hall.
- Khan, J. A., Qureshi, H. K., and Iqbal, A. (2015). Energy management in wireless sensor networks: A survey. *Computers & Electrical Engineering*, 41:159–176.
- Maia, J. E. B., Brayner, A., and Rodrigues, F. (2013). A framework for processing complex queries in wireless sensor networks. *ACM SIGAPP Applied Computing Review*, 13(2):30–41.
- Matos, T. B., Brayner, A., and Maia, J. E. B. (2010). Towards in-network data prediction in wireless sensor networks. In *Proceedings of the 2010 ACM Symposium on Applied Computing*, pages 592–596. ACM.
- Micchelli, C. A. (1984). Interpolation of scattered data: distance matrices and conditionally positive definite functions. In *Approximation theory and spline functions*, pages 143–145. Springer.
- Narayanan, R. P., Sarath, T. V., and Vineeth, V. V. (2016). Survey on motes used in wireless sensor networks: Perf & parametric anal. *Wireless Sensor Network*, 8(04):51.
- Rault, T., Bouabdallah, A., and Challal, Y. (2014). Energy efficiency in wireless sensor networks: A top-down survey. *Computer Networks*, 67:104–122.
- Rawat, P., Singh, K. D., Chaouchi, H., and Bonnin, J. M. (2014). Wireless sensor networks: a survey on recent developments and potential synergies. *The Journal of supercomputing*, 68(1):1–48.
- Durišić, M. P., Tafa, Z., Dimić, G., and Milutinović, V. (2012). A survey of military applications of wireless sensor networks. In *Embedded Computing (MECO), 2012 Mediterranean Conference on*, pages 196–199. IEEE.
- Vuran, M. C., Akan, Ö. B., and Akyildiz, I. F. (2004). Spatio-temporal correlation: theory and applications for wireless sensor networks. *Computer Networks*, 45(3):245–259.
- Xiao, H., Lu, C., and Ogai, H. (2017). A new low-power wireless sensor network for real-time bridge health diagnosis system. In *Society of Instrument and Control Engineers of Japan (SICE), 2017 56th Annual Conference of the*, pages 1565–1568. IEEE.
- Yick, J., Mukherjee, B., and Ghosal, D. (2008). Wireless sensor network survey. *Computer networks*, 52(12):2292–2330.
- Zhang, B., Liu, Y., He, J., and Zou, Z. (2013). An energy efficient sampling method through joint linear regression and compressive sensing. In *2013 Fourth International Conference on Intelligent Control and Information Processing (ICICIP)*, pages 447–450. IEEE.