

Mineração e Visualização Remota de Dados Coletados de uma Rede de Sensores Ambientais

Gerson da Penha Neto¹, N. L. Vijaykumar², Rafael Duarte Coelho dos Santos²

¹Programa de Mestrado em Computação Aplicada – CAP
Instituto Nacional de Pesquisas Espaciais – INPE

²Laboratório Associado de Computação e Matemática Aplicada – LAC
Instituto Nacional de Pesquisas Espaciais – INPE

{gerson.neto,vijay,rafael.santos}@lac.inpe.br

Abstract. *The research project Atlantic Rainforest Project Sensor Networking, implemented on the southeast coast of Brazil, is watching as a small part of the Atlantic behaves. The project's goal is to advance the understanding of the workings of the forest by collecting data detailing important phenomena such as temperature and humidity variation. In this sense it is necessary to use techniques that allow the validation of data collected even if they are not mistakenly used for research or disclosures.*

Resumo. *O projeto de pesquisa denominado Atlantic Rainforest Sensor Networking Project, implementado na costa sudeste do Brasil, está monitorando como uma pequena parte da Mata Atlântica se comporta. O objetivo do projeto é avançar no conhecimento das funções da floresta através da coleta de dados que detalham fenômenos importantes como a variação de temperatura e umidade. Neste sentido é necessário o uso de técnicas que permitam a validação dos dados coletados par que estes não sejam usados erroneamente em pesquisas ou divulgações.*

Palavras-chave: *sensoriamento remoto, mineração de dados, meteorologia.*

1. Introdução

As florestas tropicais brasileiras são de imensa importância para o mundo. Eles ajudam a regular a temperatura do clima, criar chuva, e fornecem habitat para a biodiversidade. O projeto de pesquisa denominado *Atlantic Rainforest Sensor Networking Project*, implementado na costa sudeste do Brasil, está monitorando como uma pequena parte da Mata Atlântica se comporta [Chang 2010].

Utilizando técnicas de micrometeorologia, o estudo da atmosfera logo acima do solo em escalas pequenas de tempo e espaço, fundamentais para a compreensão de como diferentes ecossistemas afetam o clima do Planeta.

A meta para os cientistas ambientais sobre o projeto foi para avançar significativamente a sua compreensão de como as funções da floresta se comportam, através da obtenção de dados mais detalhados sobre o que acontece dentro da atmosfera em uma escala pequena contida na copa da floresta.

Eles precisavam de medidas reduzidas para eventos dentro de uma área extremamente localizada: a encosta de uma colina. A escala de tempo também teve de ser comprimida, com eventos que estão sendo medidos minuto à minuto.

O projeto é uma parceria entre a Universidade de São Paulo, Fundação de Amparo à Pesquisa de São Paulo (FAPESP), Instituto Nacional de Pesquisas Espaciais (INPE), *Johns Hopkins University* e do grupo *Microsoft Research* de Pesquisa externa.

No intuito de ajudar a validar as medições feitas pelo projeto a fim de evitar que dados com falhas sejam usado em pesquisas ou em divulgações erradas o presente trabalho utilizou técnicas de mineração que podem evidenciar possíveis falhas e anomalias nos dados coletados.

2. Dados

Para se obter as medições necessárias foi desenvolvida um sistemas de sensores que poderia sobreviver as condições apresentadas por uma floresta tropical, sem considerar possíveis interferências por causa da vida animal, e transmitir dados de forma confiável.

A topologia da rede de sensores consiste de uma torre central com 60 metros de altura cercada por cinco torres menores. A torres são ligadas por cabos pendurados 2 metros acima da copa das árvores. Ao longo dos cabos foram pendurados dispositivos chamados de nós sensores, cada um reunie 3 sensores para medir temperatura e 1 para medir umidade relativa.

```
17-11-2009 23:58:38,13.72732,13.71718,14.03399,114.7875
17-11-2009 23:59:08,13.72732,13.66496,13.82586,114.7875
17-11-2009 23:59:38,13.80557,13.69107,13.79978,114.7711
18-11-2009 0:00:08,13.64896,13.61268,13.82586,114.7703
18-11-2009 0:00:38,13.70121,13.69107,13.82586,114.7481
18-11-2009 0:01:08,13.67509,13.61268,13.82586,114.6749
18-11-2009 0:01:38,13.64896,13.61268,13.79978,114.7703
18-11-2009 0:02:08,13.64896,13.58652,13.79978,114.7703
18-11-2009 0:02:38,13.64896,13.56034,13.7476,114.4669
```

Figura 1. Exemplo do arquivo ASCII de dados brutos. Cada linha contém um *timestemp*, 3 medidas de temperatura e uma medida de umidade relativa respectivamente.

Os dados processados pela *Johns Hopkins University* são devolvidos em formato *ASCII*. São arquivos de texto com extensão *.csv* contendo em cada linha um registro das temperaturas e umidade relativa obtidas dos sensores.

Cada arquivo armazena, em média, 2.880 registros por dia. Além dos registros de temperatura e umidade relativa, cada nó sensor acrescenta uma medida de tempo em cada registro, um *timestemp*, que também fica registrado no arquivo de dados brutos. As medidas obtidas por cada nó sensor são feitas a cada 30 segundos.

Neste trabalho foram utilizadas séries temporais de 53 nós sensores. Isso totaliza a quantidade de 610.560 registros diários, somando a quantidade de medições feitas pelos sensores de temperatura e umidade relativa. A Figura 1 mostra uma parte, por exemplo, de como é o arquivo de dados brutos entregue após o processamento realizada pela *Johns Hopkins University*.

3. Experimentos e Resultados

Com o propósito de validar o funcionamento dos sensores de temperatura e a veracidade das medições, utilizou-se uma rede neural SOM para agrupar as séries temporais de sensores pertencentes ao mesmo nó sensor[Kohonen 2001].

A escolha da rede SOM adveio da necessidade de formar grupos de objetos que possuem múltiplas dimensões, como é o caso das séries temporais utilizadas neste trabalho.

O agrupamento foi realizado entre as 3 séries de temperatura de cada nó sensor[SANTANA et al. 2011]. O resultado obtido foi um mapa bi-dimensional, no qual percebeu-se a formação de grupos e o distânciamento de séries que deveriam estar próximas. O resultado do agrupamento é apresentado pela Figura??.

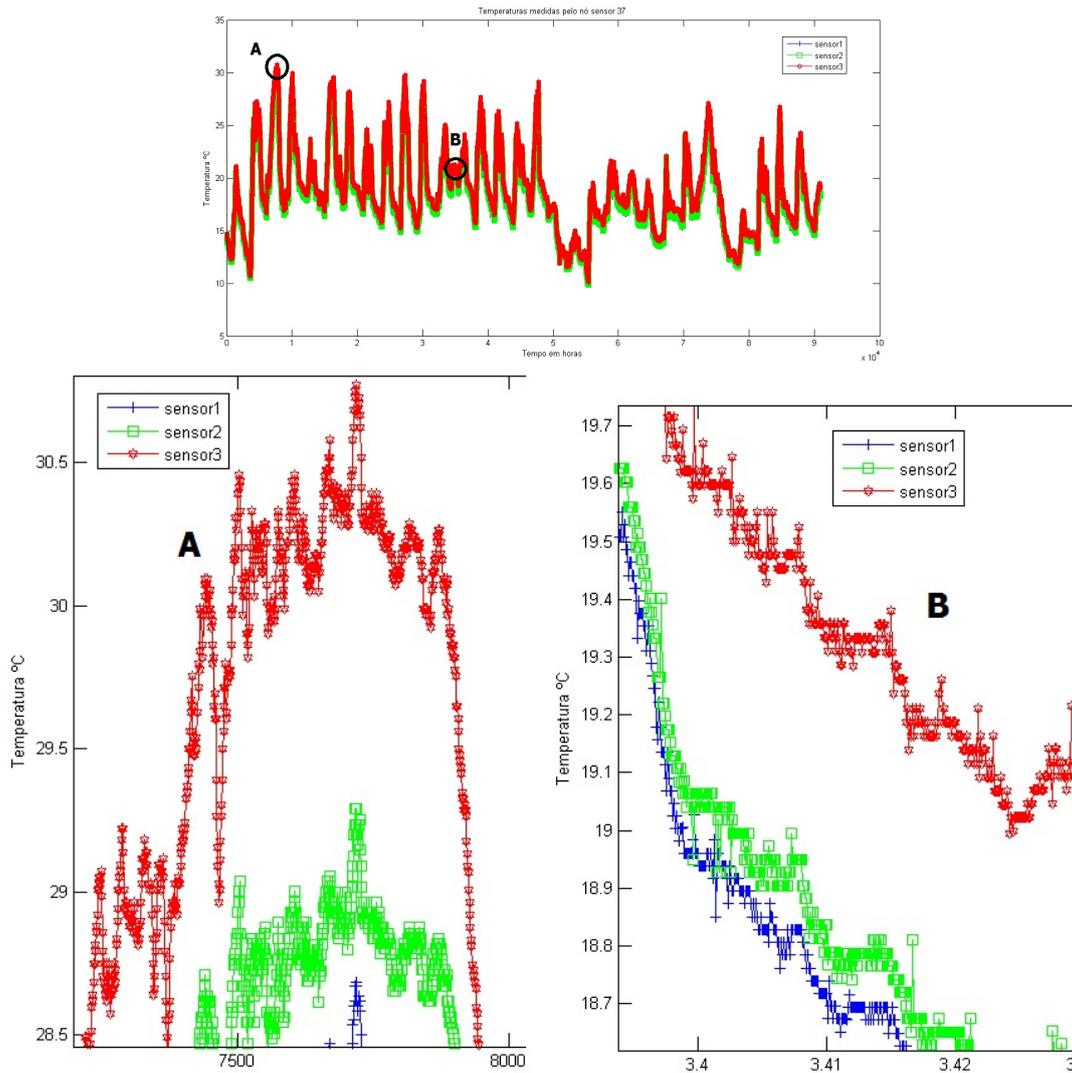


Figura 2. Anomalias encontradas no sensor 3 de temperatura do nó sensor 37

Após a fase de agrupamento foram utilizadas técnicas de representação de séries temporais para busca de anomalias, ou ruídos que possivelmente provocaram as falhas nas medições.

A principal técnica de representação utilizada foi *Symbolic Aggregate Approximation*[Lin et al. 2003]. A técnica permite a transformação de uma série temporal em cadeias de símbolos. A partir disso pode-se então usar algoritmos de busca em *Strings* para se encontrar anomalias que corresponde os erros nas medições.

Como resultado pode-se encontrar, de maneira detalhada, diferenças que provocaram o

distanciamento de determinadas séries no mapa gerado pela rede SOM. A Figura 2 mostra anomalias encontradas nas medições do nó sensor 37. Percebe-se que é difícil encontrar anomalias apenas por inspeção visual.

4. Conclusões

O presente trabalho apresentou os primeiros resultados do uso de técnicas de mineração de dados para agrupamento e busca de anomalias em séries temporais oriundas de fenômenos naturais.

O agrupamento feito pela rede SOM se mostrou eficaz no sentido de demonstrar quais séries temporais apresentavam diferenças significativas, mesmo estando geograficamente próximas.

O uso da representação por símbolos facilitou a busca por anomalias que provocaram o distanciamento entre as séries temporais no mapa gerado pela rede SOM. Isso é importante uma vez que essas anomalias são difíceis de se encontrar apenas por inspeção visual.

Evidenciou-se falhas nas séries temporais coletadas pelos sensores. Esse fato mostra que nem todos os dados devem ser considerados para o uso de pesquisas ou divulgações.

Referências

- Chang, J. (2010). Atlantic rainforest sensor networking project.
- Kohonen, T. (2001). *Self-organizing maps*. Springer, Berlin, 3 edition.
- Lin, J., Keogh, E., Lonardi, S., and Chiu, B. (2003). A symbolic representation of time series, with implications for streaming algorithms. In *Proceedings...*
- SANTANA, A. F., SANTOS, R., NETO, G. P., and VIJAYKUMAR, N. L. (2011). Aplicação de mapas auto-organizáveis para análise de dados de geosensores: Estudo de caso na mata atlântica. In *Anais...*, pages 8931–8937.