

Uma Abordagem para Identificação de Padrões de Ocorrência de Eventos Solares Transientes Baseada no Fluxo de Múons*

Mariana Teixeira¹, Daniel de Oliveira¹

¹Instituto de Computação – Universidade Federal Fluminense (IC/UFF)

teixeiramariana@id.uff.br, danielcmo@ic.uff.br

Resumo. O múon é uma partícula elementar com carga negativa, e é a componente carregada mais abundante da radiação cósmica que penetra profundamente no solo da Terra. Analisar o fluxo de múons que chega ao nosso planeta é importante para se identificar eventos solares transientes que podem causar prejuízos à população. Esse artigo apresenta uma abordagem baseada em algoritmos de mineração de sequências para automatizar o processo de identificação de padrões frequentes no fluxo de múons capturado. Experimentos realizados mostram que a abordagem é capaz de identificar eventos solares transientes por meio dos padrões frequentes identificados.

Abstract. The muon is an elementary particle with a negative charge, and is the most abundant charged component of cosmic radiation that penetrates deep into the Earth's soil. Analyzing the flow of muons that reaches our planet is important to identify solar eruptions that can cause harm to the population. This paper presents an approach based on sequence mining algorithms to automate the process of identifying frequent patterns in the captured muon stream. Experiments carried out show that the approach is capable of identifying eruptions by means of the identified frequent patterns.

1. Introdução

O Múon é a única partícula com carga elétrica capaz de penetrar profundamente no subsolo terrestre. Tal partícula se torna muito importante nas pesquisas científicas uma vez que a medição do fluxo de múons permite estudar eventos solares transientes. Os telescópios *New-Tupi* [Augusto et al. 2012] fazem a detecção de múons (aproximadamente 45.000 leituras por dia) e geram um grande volume de dados, que são atualmente armazenados em arquivos binários. O uso de arquivos é um importante limitador no que tange a execução de consultas e análises, que são fundamentais, uma vez que tais eventos podem desencadear a suspensão de diversas atividades eletromagnéticas, como por exemplo suspender as transmissões das estações de rádio.

O objetivo desse artigo é automatizar o processo de identificação de leituras do fluxo de múons [Augusto et al. 2017], por meio da aplicação de algoritmos de mineração de dados, mais especificamente de mineração de sequências [Han et al. 2011], para identificar padrões de eventos solares transientes. Para avaliar a abordagem proposta utilizamos o *dataset* dos telescópios *New-Tupi* para identificar os padrões frequentes e um segundo *dataset* fornecido pela NASA (*National Aeronautics and Space Administration*), que contém notícias relacionadas à eventos solares transientes captados (mais especificamente erupções solares), como gabarito para avaliar se os padrões identificados pelos dados dos telescópios *New-Tupi* correspondem de fato a ocorrência de eventos solares transientes.

*O trabalho aqui apresentado foi parcialmente financiado por CNPq, CAPES e FAPERJ

O artigo está organizado da seguinte forma: a Seção 2 apresenta o referencial teórico. A Seção 3 apresenta a abordagem proposta. A Seção 4 apresenta a avaliação experimental e por fim, a Seção 5 conclui esse artigo e apresenta trabalhos futuros.

2. Referencial Teórico

Esta seção tem como objetivo apresentar conceitos importantes para a compreensão da abordagem apresentada nesse artigo.

2.1. Mineração de Sequências

A Mineração de Sequências é um método de Mineração de Dados, que consiste em analisar um determinado *dataset* de sequências, onde cada sequência é composta por conjuntos de itens. Formalmente, uma sequência ∂ é representada por $\{e_1, e_2, e_3, e_n\}$, onde cada e_j , $1 \leq n$ é um evento de a sequência ∂ e e_1 ocorre antes de e_2 , que ocorre antes de e_3 e assim por diante. Existem diversos algoritmos para a mineração de sequências, como o *SPAM* [Ayres et al. 2002].

2.2. Os Telescópios *New-Tupi*

Os telescópios *New-Tupi* fazem parte de uma classe de telescópios que visa detectar múons carregados. Eles trabalham de forma sincronizada para medir continuamente o fluxo de partículas derivadas da radiação do Sol, investigando as possíveis relações entre os ciclos solares e as variações climáticas da Terra [Augusto et al. 2017].

Os telescópios *New-Tupi* são construídos a partir de quatro detectores construídos com base em um cintilador plástico (Eljen EJ-208) e uma fotomultiplicadora (Hamamatsu R877). Quando uma partícula carregada rápida (*e.g.* um múon) atravessa o cintilador, este emite uma luz fluorescente que é captada por uma fotomultiplicadora, que são sensores ópticos extremamente sensíveis a luz. A partir disso, a fotomultiplicadora converte a luz de baixa intensidade em um sinal elétrico, que é pré-amplificado até uma amplitude suficiente para posterior análise [Augusto et al. 2017]. Na lógica implementada na aquisição de dados de cada telescópio, os sinais analógicos dos detectores são digitalizados utilizando a técnica de instrumentos virtuais e as ferramentas do *software* Lab-VIEW. Ademais, cada telescópio usa um sistema anti-coincidência.

3. Abordagem Proposta

A abordagem proposta nesse artigo interage com os telescópios *New-Tupi* para identificar padrões frequentes no fluxo de múons a fim de identificar eventos solares transientes. A Figura 1 apresenta a arquitetura conceitual da abordagem proposta. O processo se inicia com o telescópio realizando as leituras do fluxo de múons e armazenando tais resultados em um arquivo binário. Ao final do dia, um *script* copia os dados para um repositório na nuvem. O componente **ETL** (do inglês *Extract Transform Load* - Figura 1 - Passo 1) acessa os arquivos produzidos e converte os dados para o formato CSV. A partir do arquivo CSV gerado, o componente de **Discretização** é ativado (Figura 1 - Passo 2). Esse componente invoca o Orange ¹, que discretiza os valores de leituras do telescópio. Uma vez que os dados se encontram discretizados o componente de **Mineração** é ativado (Figura 1 - Passo 3). Esse componente executa o algoritmo *SPAM* por meio da biblioteca SPMF ². O *SPAM* é um algoritmo que foi proposto por [Ayres et al. 2002] para minerar padrões de sequências, e que é eficiente quando os padrões de sequências são muito grandes. Uma vez que a atividade de **Mineração** é concluída, o modelo gerado já pode ser utilizado.

¹(<https://orange.biolab.si/>)

²(www.philippe-fournier-viger.com/spmf/)

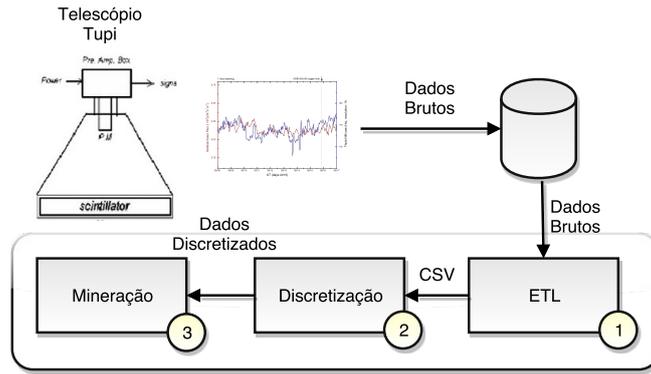


Figura 1. Arquitetura Conceitual da Abordagem Proposta

4. Avaliação Experimental

Para avaliar a abordagem proposta, uma amostra com 3.000.000 de leituras realizadas pelo telescópio *New-Tupi* foi utilizada. O *dataset* contém dados do fluxo de múons captados entre 4 de setembro e 12 de Novembro de 2014. Inicialmente, foram extraídas as *features* do repositório para serem adicionadas a um arquivo CSV, a saber: data e horário da leitura, valor *vertical* e valor *escaler*. Essa transformação foi realizada para que o Orange pudesse discretizar os dados. Os valores foram discretizados em 5 intervalos utilizando o método *Equal Frequency Discretization*. Parte do arquivo de saída é apresentado na Figura 2.

valor vertical	valor escaler	data
≥ 14.5	≥ 105.5	2014-09-04T21:00
< 8.5	94.5 - 99.5	2014-09-04T21:00
10.5 - 12.5	≥ 105.5	2014-09-04T21:00

Figura 2. Parte do arquivo de saída gerado pela discretização do Orange

Conforme apresentado na Figura 2, os valores contidos no arquivo são compostos por intervalos. Assim, houve a necessidade de transformar tais intervalos em valores únicos, pois uma estrutura sequencial é melhor minerada quando composta por valores únicos. Foi implementado um *script Python (Orange2SPMF)* que transforma o arquivo de saída gerado pelo Orange em um arquivo que contenha sequências que possam ser aceitas como entrada pelo SPMF. Para isso, os valores discretizados dos itens das sequências devem ser separados por -1, e o fim de uma sequência deve ser indicada por -2. Um fragmento do arquivo produzido pelo *script* pode ser observado na Figura 3.

2014-09-04T21:00,9 -1 100 -1 11 -1 95 -1 15 -1 95 -1 9 -1 89 -1 15 -1 95 -1 11 -1 100 -1 13 -1 100
2014-09-04T21:01,13 -1 89 -1 9 -1 95 -1 9 -1 106 -1 11 -1 106 -1 15 -1 100 -1 11 -1 89 -1 11 -1 106 (...) -2

Figura 3. Fragmento do arquivo gerado pela execução do script Python Orange2SPMF

O algoritmo *SPAM* foi executado consumindo o arquivo anteriormente citado, explorando diversas combinações dos parâmetros *minsup* (suporte) e *Min Pattern Length* (quantidade mínima de elementos nos padrões encontrados) porém com *Max Gap* com valor fixo 1, que não permite que haja intervalos entre os elementos. Os diversos arquivos de saída do SPMF com os padrões encontrados foram integrados em um único arquivo.

Após a mineração, foi necessário encontrar em quais datas ocorreram cada padrão descrito no arquivo com as saídas agregadas. Essas novas informações obtidas foram consolidadas em um novo arquivo, que é composto por uma coluna relacionada ao padrão

encontrado, uma coluna relacionada ao suporte do padrão e outra coluna relacionada às datas em que esse padrão ocorreu.

Para avaliar a abordagem proposta e validar os resultados obtidos com a mineração de sequências, verificamos, utilizando um *dataset* fornecido pela NASA (*National Aeronautics and Space Administration*), se nas datas em que os eventos frequentes foram identificados de fato ocorreram erupções solares. A NASA possui um site³ que contém o arquivo referente às notícias relacionadas à eventos solares que ocorreram entre 2010 e 2015. De posse das datas e dos padrões frequentes, foi possível analisar no site da NASA os dias e classificações dos eventos solares transientes que ocorreram na mesma faixa de tempo dos registros coletados pelo telescópio *New-Tupi*. Com isso, foram cruzados os dados do arquivo de padrões com as datas de eventos solares identificados pela NASA, sendo assim possível criar um novo arquivo que contém os padrões, contidos nas análises de múons feitas pelo *New-Tupi* nas datas de eventos solares noticiados pela NASA. Como muitas sequências foram encontradas, foi necessário minerar novamente o este último arquivo visando descobrir qual o padrão de fluxo de múons que pode determinar os eventos solares noticiados pela NASA. Novamente, o *SPMF* foi utilizado para minerar tais sequências. O arquivo de saída pode ser observado na Figura 4. Foram identificados eventos solares em 10/09/2014, 19/10/2014 e 30/10/2014, o que coincide com o *dataset* fornecido pela NASA.

```
9 -1 89 -1 9 -1 89 -1 #SUP: 32
89 -1 9 -1 89 -1 9 -1 #SUP: 29
```

Figura 4. Saída gerada pelo SPMF

5. Conclusão e Trabalhos Futuros

Os telescópios *New-Tupi* realizam leituras dos fluxos de Múons que chegam ao nosso planeta. A abordagem proposta nesse artigo visa encontrar padrões nos valores referentes ao fluxo de múons capturados pelos telescópios *New-Tupi*, baseando-se no algoritmo de mineração de sequências *SPAM*. De forma a validar os resultados experimentais obtidos, cada padrão frequente identificado foi comparado com um *dataset* da NASA que descreve eventos solares identificadas pela instituição. Todas as sequências identificadas coincidem com eventos identificados pela NASA, mostrando que a aplicação da abordagem proposta é válida e promissora. Apesar de representar um avanço, novos experimentos de maior escala se fazem necessários com o *dataset* completo dos telescópios *New-Tupi*.

Referências

- Augusto, C. R. A., Kopenkin, V., Navia, C. E., Tsui, K. H., and Sinzi, T. (2012). Search for a simultaneous signal from small transient events in the pierre auger observatory and the tupi muon telescopes. *Phys. Rev. D*, 86:022001.
- Augusto, C. R. A., Navia, C. E., de Oliveira, M. N., Nepomuceno, A. A., Kopenkin, V., and Sinzi, T. (2017). Muon excess at sea level during the progress of a geomagnetic storm and high-speed stream impact near the time of earth's heliospheric sheet crossing. *Solar Physics*, 292(8):107.
- Ayres, J., Flannick, J., Gehrke, J., and Yiu, T. (2002). Sequential pattern mining using a bitmap representation. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 429–435. ACM.
- Han, J., Pei, J., and Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.

³https://www.nasa.gov/mission_pages/sunearth/news/solar-event-2010-2015