

Protótipo de Sistema Web para Auxiliar na Mineração de Dados Temporais

André Gustavo Maletzke¹, Leandro Borges dos Santos¹, Richardson Floriani Voltolini¹, Hwei Diana Lee¹, Feng Chung Wu^{1,2}

¹Laboratório de Bioinformática – LABI, UNIOESTE, Foz do Iguaçu, PR

²Serviço de Coloproctologia, DMAD, FCM, UNICAMP, Campinas, SP
{sborgesleandro, andregustavom}@gmail.com

Objetivos

Processos que visam extrair conhecimento de conjunto de dados, como o de Mineração de Dados (MD), apoiado por técnicas de Aprendizado de Máquina, quando aplicados diretamente em Séries Temporais (ST), desconsideram o fator temporal presente nesses dados. Sob esse escopo, este trabalho tem como objetivo apresentar um protótipo *Web* para auxiliar na MD temporais, mediante a identificação de *motifs*.

Métodos/Procedimentos

O desenvolvimento do protótipo foi realizado em quatro etapas. Na Etapa 1, foram definidos os requisitos básicos. A Etapa 2 foi destinada ao projeto da solução e na Etapa 3 foram definidas as tecnologias e métodos utilizados, sendo eles: linguagem de programação *Ruby* e o *framework Ruby on Rails*, bem como as seguintes tecnologias para aplicações *Web*: *JavaScript*, *HyperText Markup Language (HTML)*, *Cascading Style Sheets (CSS)* e *Asynchronous Javascript and XML (AJAX)*. Para permitir o acesso a aplicação foi utilizado o servidor *WEBrick* e o banco de dados *MySQL*. Para realizar a identificação de *motifs* foram utilizadas a abordagem por Força Bruta (FB) [1], e a abordagem probabilística denominada *Random Projections (RP)* [2]. Na implementação dos algoritmos de identificação de *motifs* foi utilizada a linguagem *R* e para o desenvolvimento do protótipo utilizou-se o método de Prototipação. Na Etapa 4, para avaliar o protótipo foi realizado o seguinte procedimento: selecionou-se a ST de Eletrocardiograma (ECG) de Koski de domínio público, onde foram extraídos as 1.000 primeiras observações e foi realizada a inserção de *motifs* artificiais com 250 observações. Após, foi utilizado o protótipo no intuito de verificar a

eficiência na identificação de *motifs*. Esse processo foi repetido dez vezes.

Resultados

Os requisitos definidos foram cumpridos, dentre eles a visualização gráfica dos resultados, conforme Figura 1.

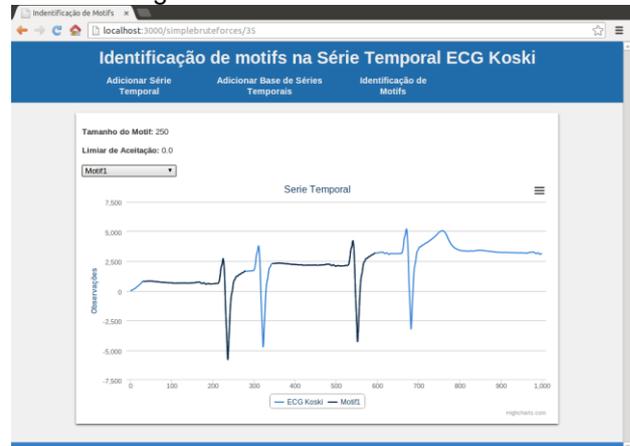


Figura 1: Tela de identificação de *motifs*

O resultado das simulações realizadas foram satisfatórios, pois apresentaram tanto para a abordagem de FB quanto para a RP uma precisão de 100%, isto é, todos os *motifs* foram identificados por ambas abordagens.

Conclusões

Com desenvolvimento deste protótipo espera-se auxiliar na mineração de dados temporais mediante a identificação de padrões morfológicos.

Referências Bibliográficas

- [1] LIN, J.; KEOGH, E.; LONARDI, S.; PATEL, P. Finding motifs in time series. VIII SIGKDD, 2002.
- [2] CHIU, B.; KEOGH, E.; LONARDI, S. Probabilistic discovery of time series motifs. IX SIGKDD, 2003.