



Universiteit
Leiden
The Netherlands

Methods and tools for mining multivariate time series

De Gouveia da Costa Cachucho, R.E.

Citation

De Gouveia da Costa Cachucho, R. E. (2018, December 10). *Methods and tools for mining multivariate time series*. Retrieved from <https://hdl.handle.net/1887/67130>

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/67130>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The following handle holds various files of this Leiden University dissertation:

<http://hdl.handle.net/1887/67130>

Author: de Gouveia da Costa Cachucho, R.E.

Title: Methods and tools for mining multivariate time series

Issue Date: 2018-12-10

Resumo

O tempo é uma dimensão fundamental para a explicação de muitos fenómenos. A situação económica de um determinado país, o comportamento expectável de um cliente, ou o diagnóstico das causas de uma doença são só alguns exemplos de fenómenos que podem ser melhor entendidos com dados ordenados temporalmente (séries temporais).

Com a consolidação da era digital, a recolha e armazenamento de series temporais tornou-se omnipresente. Um correcta interpretação de series temporais possibilita uma melhor compreensão de fenómenos actuais, a extração de tendências e a previsão de eventos futuros. Agora, num mundo cada vez mais conectado através de sensores e monitorização de comportamentos digitais, a extração de conhecimento de séries temporais poderá tornar-se uma grande desafio. Foi este desafio que tornaram os métodos de aprendizagem automática (normalmente denominados de inteligência artificial) em ferramentas essenciais nos dias que correm.

A extração de conhecimento de séries temporais é um campo especializado num tipo de estrutura de dados onde, para cada sistema ou sujeito, as variáveis são medidas várias vezes ao longo de um determinado período de tempo. Nesta tese focamos nesta estrutura de dados e consideramos o caso de séries temporais multivariadas. Na maioria dos casos, consideramos também que as variáveis são medidas a diferentes frequências (uma vez por cada dia, hora, minuto, segundo...). Por exemplo, variáveis poderão ser indicadores de actividade física (número de passos, número de patamares subidos a pé, ...), registos médicos (historial médico, variáveis metabólicas, ...) e indicadores económicos (proprietário de imobiliário, salário bruto, ...). Quando combinadas, estas variáveis podem ser exploradas com múltiplos fins, através de métodos de inteligência artificial.

Nesta tese de doutoramento, propomos métodos supervisionados e não supervisionados de aprendizagem automática (algoritmos), capazes de lidar com séries temporais. Primeiro, propomos um método de análise de clusters (bi-

clustering). Este, foi capaz de encontrar conjuntos de variáveis que mostram um comportamento similar e consistente num determinado número de segmentos temporais. Alternativamente, numa situação supervisionada, dada uma variável dependente e um conjunto de variáveis independentes, propomos um outro método que transforma e agrega as variáveis independentes de forma inteligente. Os resultados experimentais apresentam o potencial de ambos os métodos acima referidos.

Para além dos métodos de aprendizagem automática, nesta tese introduzimos duas ferramentas de Software enquanto serviço (*Software as a Service*). Nestas ferramentas, utilizadores sem conhecimentos de programação podem seguir de forma intuitiva, o processo de aprendizagem automática e teste dos métodos mencionados no parágrafo anterior. Estas ferramentas permitem efectuar todo o processo de um cientista de dados, começando pelo processo de tratamento de dados, aprendizagem, optimização e avaliação de modelos. A principal contribuição destas ferramentas não são os métodos de inteligência artificial, mas sim o apoio aos utilizadores destes métodos.

Por fim, apresentamos um método de inteligência artificial aplicado a um desporto de alta competição. O objectivo deste estudo é a melhoria dos resultados nas competições de atletas de patinagem de velocidade, um desporto com uma grande dimensão económica e social na Holanda. Mais concretamente, utilizando os dados de treino de uma equipa de alta competição, construímos modelos que optimizam os resultados competitivos através de ajustamento nos treinos. O nosso estudo mostra que, através de modelos personalizados, podemos evitar o treino excessivo ou deficiente.