

Abstract

This thesis applies the Signature method as a measurement of similarities between two time-series objects, using the Signature properties of order 2, and its application to Asymmetric Spectral Clustering. The method is compared with a more Traditional Clustering approach where similarities are measured using Dynamic Time Warping, developed to work with time-series data. The intention for this is to consider the traditional approach as a benchmark and compare it to the Signature method through computation times, performance, and applications. These methods are applied to a financial time series data set of Mutual Exchange Funds from Luxembourg.

According to Chevyrev and Kormilitzin (2016), the Signature transformation converts the data into multi-dimensional paths using various embedding algorithms and then proceeds to the computation of individual terms of the Signature which summarise certain information contained in the data. Levin, Lyons, and Ni (2013) introduced the possibility of its use to understand financial data and the potential this approach has for machine learning and prediction.

In our work, the approach is applied to clustering, where the goal is to build homogeneous groups of observations and discover hidden patterns in data. Performing clustering evolves at least three steps: (1) selecting a measure able to quantify the similarity between objects; (2) choosing the right method to perform clustering, according to the specific problem; (3) setting the number of desired clusters. The order of the steps depends on the algorithm chosen to cluster the data.

After the literature review, we introduce the Dynamic Time Warping method and the Signature method. We continue with the explanation of Traditional Clustering approaches, namely k-Means, and Asymmetric Clustering techniques, namely the k-Axes algorithm, developed by Atev (2011). The last chapter is dedicated to the Practical Research where the previous methods are applied to the data set. Results confirm that the Signature method has indeed potential for machine learning and prediction, as suggested by Levin, Lyons, and Ni (2013).

Key words: Signature, Dynamic Time Warping, Clustering, Machine Learning.

Resumo

Esta dissertação aplica o método da Signature como medida de similaridade entre dois objetos de séries temporais usando as propriedades de ordem 2 da Signature e aplicando-as a um método de Clustering Assimétrico. O método é comparado com uma abordagem de Clustering mais tradicional, onde a similaridade é medida usando Dynamic Time Warping, desenvolvido para trabalhar com séries temporais. O intuito é considerar a abordagem tradicional como benchmark e compará-la ao método da Signature através do tempo de computação, desempenho e algumas aplicações. Estes métodos são aplicados num conjunto de dados de séries temporais financeiras de Fundos Mútuos do Luxemburgo.

De acordo com Chevyrev e Kormilitzin (2016), a transformação da Signature consiste em converter os dados em caminhos multidimensionais por meio de vários algoritmos de incorporação e, em seguida, procede ao cálculo dos termos individuais da Signature que resumem certas informações contidas nos dados. Levin, Lyons e Ni (2013) introduziram a possibilidade do seu uso para entender dados financeiros e o potencial que esta abordagem tem para Machine Learning e Previsão.

Esta abordagem é aplicada a clustering, onde o objetivo é criar grupos homogêneos de observações e descobrir padrões escondidos nos dados.

A execução de clustering envolve pelo menos três etapas: (1) selecionar uma medida capaz de quantificar a similaridade entre os objetos; (2) escolher o método certo para realizar o clustering, de acordo com o problema específico; (3) definir o número de clusters desejados. A ordem das etapas depende do algoritmo escolhido para agrupar os dados.

Após a revisão da literatura, apresentamos o método Dynamic Time Warping e o método da Signature. Prossegue-se com a explicação das abordagens de Clustering Tradicional, nomeadamente k-Means, e Clustering Espectral Assimétrico, nomeadamente k-Axes, desenvolvido por Atev (2011). O último capítulo é dedicado à Investigação Prática onde os métodos anteriores são aplicados ao conjunto de dados. Os resultados confirmam que o método da Signature têm efectivamente potencial para machine learning e previsão, como sugerido por Levin, Lyons and Ni (2013).

Key words: Signature, Dynamic Time Warping, Clustering, Machine Learning.