# Abstract

Time series (TS) is an integral part of the modern life that helps to understand the several aspects of the daily activities. For example, the trend in weather time series gives an idea about the climate conditions of a place, the trend in infectious disease time series data gives an idea about rate at which the disease rises and the preventive measures to be taken. Several data collection devices like sensors, data loggers, data streaming platforms are used to accumulate the time series data from multiple domain. TS analysis on the historical data provide a basis to derive the underlying patterns and anomalies. It further helps in evaluation and deciding the performance of the systems, models and identify the areas that require improvements.

Several research questions arise for time series data mining operation and this thesis intends to address the following 1. how to preserve the pattern based information when using the dimensionality reduction of the TS data representation techniques 2. how to identify the frequently occurring temporal patterns from the historic time series datasets at the same or different time granularity 3. how to identify the atypical patterns from the historic time series datasets 4. out of the many time series samples present, how to identify the time series samples that have a regularity in patterns.

This thesis aim to contribute towards the above discussed research questions in case of TS datasets. It is also important to consider the space and time constraints that will be required to mine each of the patterns present in the entire TS dataset. In this regard, the first study of this thesis presents a TS dataset representation technique that can preserve the pattern based information. A weighted directed graph structure is formed from a dimensionaly reduced TS dataset. A graph based clustering approach is proposed that group the TS samples based on similar temporally dependent subsequent patterns. The graph path intend to capture the temporal patterns observed in TS dataset. The frequently occurring longest temporal patterns from each cluster is identified. The temporally dependent atypical patterns are extracted from the TS dataset using the simple graph component analysis. The next study of this thesis propose a TS subsequence clustering approach in case of TS dataset. The proposed clustering technique can identify the dense coherent substructures which actually represent the subsequently co-occurring patterns of the time series samples. The cluster labels are further used to compute the regularity of TS samples. Regularity denote how frequently does the cluster labels change in each TS sample. The temporally dependent atypical patterns are extracted using the vertex degree analysis on the graph structure. The third contributory work of this thesis propose a multiple graph structure formation to capture the subsequently co-occurring patterns at multiple time granularity in TS dataset. An automated feature extraction from the subsequences is shown using state-of-the-art model. The cluster labels are obtained using quasi-clique based subsequence clustering approach applied at each granular level. The cluster labels are further used to compute the regularity of TS samples at each granular level. It is shown that which of the subsequent patterns are common at different granular level and which of them are different. The final contributory work of this thesis is to classify the TS samples based on the regularity in patterns. Unlike the previous work where a graph structure is designed, in this work, the graph structure is formed automatically by learning which of the previous subsequences are important to predict the future subsequence. The attention values obtained from the state-of-the-art model define the importance of the subsequences when predicting the future subsequence values, that are used for edge formation. The learnt graph structure is used to train the Graph convolutional network (GCN) based classifier model.

The clustering, classification techniques employed to solve the above research questions is useful in the fields of engineering, finance and utility domain. For example, in case of stock market TS data, the opening and the closing prices of a stock differ, and also there exist several companies with similar rise/fall in patterns in their opening or closing stock prices. Clustering and identification of such companies help in designing accurate recommender systems. Study of the frequently occurring temporal patterns in case of stock market help in identifying the companies which have similar patterns during a certain range of time. Identifying the companies that follow similar rise/falls in pattern at the different time granularity help in identifying the differences or the similarity between the given time granularities and study the change in market.

Considering the smart grids utility, the historic TS dataset consist of electricity power consumption for each residential/commercial building. Considering each building as a consumer, shows different behavioral usage of appliances that cause pattern changes. Clustering such consumers help the electricity utility providers in decision making. Study of the frequently occurring temporal patterns in case of smart grids, it help to identify the appliances, devices that are dependent in nature, or are used subsequently. The DSM is the current state-of-the art decision making system in smart grids that help in maintaining a demand-supply equilibrium. The regularity measure as shown in this thesis is important choosing the appropriate consumers for DSM operations. Identifying the TS samples that have temporal patterns that are different from the majority other consumers is important and that gives idea about theft, both in case of the stock market and smart grids. An automated process, like a classifier, that can automatically find the consumers that are suitable for DSM operations, will reduce the workload of electricity utility providers.