UROP1100 Final Report

Title: Irregular Sampled Time Series

Abstract:

Time series based data has seen

Over the past decade

Plenty of previous models

Over the past decade

Both statistical and machine learning based approaches have been successfully used to map

Time series data has

Irregularly sampled time series (ISTS) data is categorized by non-uniform intervals between observations. This structural property makes traditional machine learning and statistical approaches of modeling time series data insufficient.

Statistical and machine learning models have been applied to learn from time series data

Time series data has been a focus of intensive research, and several machine learning and statistical models have been defined to analyze and forecast said series. However, these models assume uniform intervals between observations, and hence require modifications to the underlying structure to account for time lag between samples in case of Irregularly sampled time series (ISTS) data. This irregularity is more adamant in multivariate time series where multiple variable can each have different sampling rates. Being able to predict and classify using ISTS data is important in several domains such as health, patient vitals measured depending on symptoms, and astronomy where sampling rate can be affected by various factors for instance the weather conditions. This report covers recent state-of-the-art deep learning based approaches for such data.

We start by desciribing ISTS data and its properties. As the name implies, ISTS consists observations taken over a certain time interval. This is

We start by describing various applications of this data. The main two uses of ISTS data are in forecasting future values which can be helpful in predicting weather situation or risk of typhoons and help take proper measures, and classification where it is used to predict the likelihood of a certain disease or mortality rate of a patient. Other uses such as interpolation and detection also exist.

Irregularly sampled time series (ISTS) data arise naturally in several real world applications

Irregularly sampled time series (ISTS) data appears in several real world applications including but not limited to health, economy, science and astrology.

Introduction:

Time series data has been a focus of intensive research and several machine learning and statistic models have been designed to analyze and forecast using said data. However, these models assume uniform intervals between observations, and hence require modifications to the underlying structure to account for time lag between samples in case of Irregularly sampled time series (ISTS) data. The irregularity here refers to both the intra-series irregularity resulting from irregular sampling interval and inter-series irregularity in multivariate time series mainly caused by their varying sampling rates. Being able to predict and classify using ISTS data is of immense importance in several domains such as economy, astrology, science and health where ISTS data arises naturally. This report summarizes recent deep learning based approaches used to analyze ISTS data with a main focus on their uses in health sector.

Intro Part II

The irregularity in ISTS data can be both intra-series, where samples are taken at irregular intervals, and inter-series as different time series can have measurements taken at different sampling rates. To illustrate these relations, we consider patient monitoring at hospitals. The vitals of patient such as heart rate are sampled at a higher rate as compared to for instance blood glucose level. This leads to inter-series irregularity between time series of heart rate and blood glucose level. Furthermore, measurement of blood glucose level will depend on various other factors such as patient’s symptoms or even heart rate. This amounts to intra-series irregularity.

Another key point highlighted in the above scenario is the dependence of two series on one another. As multiple series might intermingle and depend on one another, it is necessary that whatever model we describe not only takes into consideration intra-series local trend, but looks at the global picture of all the series data available. Next, we look into possible results we can obtain through the analysis of ISTS data.

Much like normal time series data, ISTS can be used for data forecasting where all previous records before time t\* are available, and the objective is to find the record at time (t\* + delta) where delta > 0. Another common use is in classification problems. In the patient monitoring system, ISTS data can be used classify different diseases or mortality rate of patients. Interpolation using ISTS data allows the missing points in the series to be estimated. This can be done either using only the data available before the point being interpolated (detection) or more often using entire series both the future and past values to give better estimate for present value (interpolation). We will be focusing on forecasting and classification problems in this report although interpolation will be used at initial stages to prepare data for use in models.

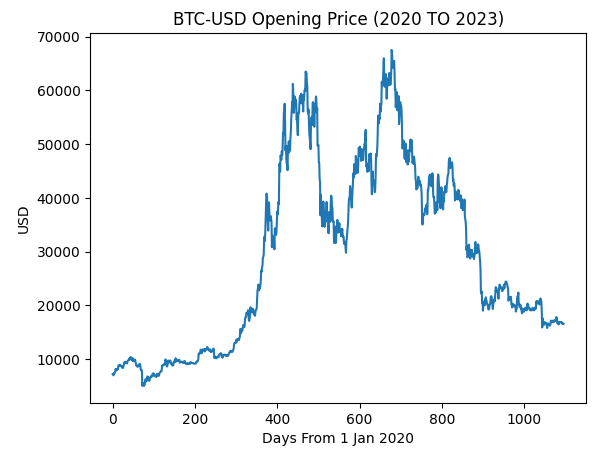
We breakdown our problem of ISTS data into 2 subproblems, dealing with data irregularity and using the multiple series collectively to analyze and predict meaningful results. If we can remove the irregularity from our ISTS problem, we can apply the standard techniques used on time series for our purposes. These techniques include statistical models such as ARIMA where the series is segregated into components and trends are analyzed or Vector AutoRegression which is ARIMA for multivariate time series. While these models are easier to train and give reasonably good results, we will focus only on the machine learning based approaches as these will later be used in solving ISTS. The simplest machine learning based approach is to use polynomial regression on the time series. This normally does not yield good result as the underlying time series usually can not be represented by a polynomial equation. To improve this, we use the concept of windowing. By taking small windows over the entire series, we try to map the smaller window

TAKE 2

Modeling raw data in the form of large text corpus, images, or sequences into much more useful results like sentiments, prices, and weather has seen much development over the last decade. One such form of raw data used to forecast future trends is time series. A time series is categorized by having an additional dimension namely time of observation. Readings taken are implicitly related to the time of observation and their order itself carries valuable information. This makes forecasting or classification using said data challenging for traditional machine learning models where the order of readings was insignificant. This report will briefly look into various techniques used to model time series data and then move into its main part: data classification and forecasting using Irregularly Sampled Time Series (ISTS) data. In this report, we will mostly discuss deep learning-based approaches and discuss recent works in the fields. Lastly, we will present our modification of Q-Network architecture and compare its results with other models on the MMIC II dataset.

The code is found at:

Introduction:



Time series data adds a temporal dimension into our regular data. As shown in figure 1, this makes the data irregular and linear functions are unable to fit this model. Statistical approaches such as ARIMA work by segregating the data into the Trend, Seasonality, Cyclic and Random components. However, ARIMA requires the provided data to be stationary i.e. its statistical properties like mean, variance and covariance is independent of time. A non-stationary series can be converted into stationary by means of segregation as described above, but using this approach has its limits. Moreover, the data is seldom uni-variate and often comes as a collection of multiple time series which all include useful information for making predictions. While other statistical model such as VARMA extend the capability of ARIMA to multivariate data analysis, we will look into machine and deep learning based approaches only.

The data we will be using for comparing efficiency of our models is BitCoin to USD price from 1 Jan 2020-23. The aim of our model will be to forecast future opening price of the for BTC-USD given the past opening, closing, highest and lowest prices. For evaluation, we will use our model to predict the opening price over the first 6 months of 2023 and use use RMSE as error metric. In the next section we will compare three different models to map this data namely regression, CNN, and LSTM.