International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

Extended Abstract
Vol 2, Iss 4
2019

Machine Learning 2019: Time Series Anomaly Detection using CNN - coupled with data augmentation using GANs - Prasenjeet Acharjee - Baghmane World Technology Park, - India

Prasenjeet Acharjee
Baghmane World Technology Park, - India

Standardized data sets have been a crucial factor in the success of ML. There are a lot of standard datasets such as MNIST database of handwritten digits, ImageNet database etc. which have been used in other domains to great success to drive adoption of ML. However, there is nothing like that available for the telecom domain in addition to other issues such as many/most data sets are toy, noisy, unnormalized; some data sets are proprietary, most of the data is non-iid [independent and identically distributed random variables]. The most obvious drawback being that most of our data sources for telecom are representations of network data that were not built for ML.

To label any historical telecom data, a network domain expert needs to manually label the anomaly from normal network behavior and then label the different types of anomalies in total population of anomaly data to apply a supervised learning algorithm on it to classify various cell issues. However, due to the high volume, veracity, dimensionality, and cardinality of input data, manual analysis of the massive amounts of data and associated metrics is inefficient and practically not feasible. Manual analysis is also not sustainable as this method is subject to individual personnel knowledge and experience which can result in inconsistencies and rendering the process non-scalable. Due to the proceeded digitization of mechanical and societal forms, counting the sending of organized sensors, we are seeing a fast multiplication of time-ordered perceptions, known as time arrangement. For case, the behavior of drivers can be captured by GPS or accelerometer as a time arrangement of speeds, bearings, and increasing speeds. We propose a system for exception discovery in time arrangement that, for illustration, can be utilized for distinguishing unsafe driving behavior and perilous street areas. Particularly, we to begin with propose a strategy that produces factual highlights to enhance the highlight space of crude time arrangement. Following, we utilize an autoencoder to reproduce the enhanced time arrangement. The autoencoder performs dimensionality diminishment to capture, employing a small highlight space, the foremost agent highlights of the improved time arrangement. As a result, the reproduced time arrangement as it were capture agent highlights, while exceptions frequently have non-representative highlights.In arrange to distinguish exceptions in hydrological time arrangement information for progressing information quality and decision-making quality related to plan, operation, and administration of water assets, this investigate creates a time arrangement exception discovery strategy for hydrologic information that can be utilized to recognize information that go astray from authentic designs. The strategy to begin with built a estimating demonstrate on the history information and after that utilized it to anticipate future values. Inconsistencies are accepted to require put on the off chance that the watched values drop exterior a given expectation certainty interim (PCI), which can be calculated by the anticipated esteem and certainty coefficient. The utilize of PCI as edge is mainly on the truth that it considers the vulnerability within the information arrangement parameters within the forecasting model to address the reasonable edge choice issue. The strategy performs quick, incremental assessment of information because it gets to be accessible, scales to huge amounts of information, and requires no preclassification of peculiarities.Results of such circumstances in hydrological data frameworks may result within the DRQP (information wealthy, but quality poor) wonder. Subsequently, the first checking information (i.e., precipitation, release, and water levels) ought to experience a preprocessing step to kill the negative impact caused by erroneous or irregular information due to instrumented flaws, information inborn alter, operation blunder, or other conceivable

affecting components . ce, exception discovery, as a rule, gets to be an imperative step for hydrologic time arrangement examination based on the checking information. With respect to little checking datasets, information supervisors can distinguish and bargain with exceptions straightforwardly with a basic graphical or manual handle. This ponder creates a real-time exception location strategy that utilizes a window-based estimating show for hydrologic time arrangement collected from programmed checking frameworks. The strategy builds a determining show from an arrangement of chronicled point values with a given window to foresee future values. In case the watched esteem contrasts from the anticipated esteem past a certain limit, an exception would be shown. The strategy employments expectation certainty interim () as an edge in the thought of instability within the data arrangement parameters within the estimating demonstrate. Information is classified as anomalous/nonanomalous based on whether or not they drop the exterior a given . In this way, the strategy gives a principled system for selecting a limit. This strategy does not require any pre-classified illustrations of information, scales well to huge volumes of information, and allows for quick incremental assessment of information because it gets to be accessible.In arrange to assess the proposed strategy, it was connected to two distinctive hydrological factors, water level and everyday stream, from Huayuankou (in the future , 34.76°N, 113.58°E) and Lanzhou (in the future , 36.04°N, 103.49°E) stations gotten from national hydrology database of MWR, China. The comes about to appear that the proposed strategy can exactly detect the exceptions within the hydrological time arrangement with close insignificant untrue positive rate. Besides, the algorithm's proficiency is analyzed based on the location results. The rest of the paper is organized as takes after. Within the another area (Area 2) we show the related work to this region of inquire about. In Area 3 we display subtle elements around the proposed calculation for exception location in time arrangement based on expectation certainty interim. A number of tests with the proposed strategy utilizing real-world hydrological time arrangement are detailed in Segment 4. At last, Segment 5 gives conclusions and proposals for advance inquire about.

As the fundamental resources for water resources management and planning, long-term hydrological data are sets of discrete record values of hydrological elements that are collected with time and have been frequently analyzed in the field such as flood and drought control, water resources management, and water environment protection. With the development of data acquisition technology and data transmission technology, hydrological departments collected ever-increasing amounts of time series data from automatic monitoring systems via loggers and telemetry systems. Within these datasets, hydrologic time series analysis becomes workable and credible for building mathematical model to generate synthetic hydrologic records, to forecast hydrologic events, to detect trends and shifts in hydrologic records, and to fill in missing data and extend records. However, a hydrologic time series is generally composed of a stochastic component superimposed on a deterministic component and usually shows stochastic, fuzzy, nonlinear, nonstationary, and multitemporal scale characteristics

In this talk, I intend to cover brief overview of how time series anomaly detection problems can be tackled with an unconventional approach which captures the multi-spatial relationships between time and various features, augment the dataset using deep generative models (GANs) and train a classifier to achieve state of the art results. I also intend to briefly touch upon an evaluation framework for measuring GAN performance by evaluating on explicitly parameterized, synthetic data distributions that can be applied to any dataset.