1

Introduction

1.1 Background

“An outlier is an observation which deviates so much from the other observations as to arouse suspicions that it was generated by a different mechanism.” D. Hawkins. Identification of Outliers, Chapman and Hall, 1980.

Applications

Sensor data, mechanical systems diagnosis, medical data, network intrusion data, newswire posts or financial posts

Components of time series

Types of anomalies

Contextual - values at specific time stamps suddenly change with respect to their temporally adjacent values.

Collective - entire time series or large subsequences within a time series have unusual shapes [Aggarwal, 2013]

Aggarwal

9.2 Prediction-Based Outlier Detection in Streaming Time Series

For contextual anomalies. The most common application of temporal outlier detection is that of detecting deviation-based outliers of specific time-instants with the use of regression-based forecasting models. (p.276)

Correlations across time: This is the same principle as temporal continuity, which is typically implemented using autoregressive modeling and forecasting. Significant deviations from the expected (i.e., forecasted) predictions are defined as outliers. Such significant deviations are therefore defined by violations of temporal continuity. (p.276)

Correlations across series: Many sensor applications result in time series that are often closely correlated with one another. For example, a bird call at one sensor will typically also be recorded by a nearby sensor. In such cases, one series can frequently be used in order to predict another. Deviations from such expected predictions can be reported as outliers. (p.276)

Autoregressive models, with focus on multivariate models in 9.2.2. These are:
* VARIMA (9.2.2.1, p.279), can be slow
* PCA-based techniques (9.2.2.3, p. 282), generally more robust to the presence of noise and outliers. Most well-known is SPIRIT.

9.3 Time-Series of Unusual Shapes

For collective anomalies.

Full-series anomaly: In this case, the shape of the entire series is treated as an anomaly. This shape is compared against a database of similar time series. However, in most cases, unless the database of sequences corresponds to a relatively short segment of time-stamps, the noise variations within the series will mask the anomalous shape. This is analogous to the problems of noise encountered in detecting outliers in high- dimensional data. (p.287)

Subsequence-based anomaly: If the time series is collected over long periods of time, then we might have a single time series that shows typical trends over shorter time-periods. Therefore, the anomalous shape is detected over small windows of the time series as deviations from these typical trends. (p.287

Methods: Transformation to Other Representations (9.3.1) Distance-Based Methods (9.3.2) Probabilistic Models (9.3.3) Linear Models (9.3.4)

Vertical analysis In time-series analysis, vertical analysis is more important where each individual series (or dimension) is treated as a unit, and the analysis is primarily performed on this unit. In the event that multiple series are available, cross-correlations may be leveraged, although they typically play a secondary role to the analysis of each individual series. This is because time-series data is contextual, which imposes strong temporal locality on series values. (p.273)

Labels Labels may be available to supervise the anomaly detection process in the time-series detection settings. In the time-series setting, the labels may be associated with time-instants, with time intervals, or they may be associated with the entire series. (p.275)

Supervised vs. unsupervised In general, supervised methods almost always perform better than unsupervised methods because of their ability to discover application-specific abnormalities. The general recommendation is to always use supervision when it is available. (p.275)

Generally, unsupervised methods can be used either for noise removal or anomaly detection, and supervised methods are designed for application-specific anomaly detection. Unsupervised methods are often used in an exploratory setting, where the discovered outliers are provided to the analyst for further examination of their application-specific importance. (p.4)

Noise In the unsupervised scenario, where previous examples of interesting anomalies are not available, the noise represents the semantic boundary between normal data and true anomalies – noise is often modeled as a weak form of outliers that does not always meet the strong criteria necessary for a data point to be considered interesting or anomalous enough. (p.3)

Relationship between Unsupervised Outlier Detection and Prediction The methods in this section use supervised prediction and forecasting methods for unsupervised outlier detection. Outliers are, after all, violations of the “normal” model of data dependencies. A prediction model, therefore, helps in modeling these dependencies as they apply to a specific data point. Violations of these dependencies represent violation of the model of normal data and therefore correspond to outliers. (p.284)

Questions to be answered: How could one classify different types of anomalies? How to distinguish between noise and anomalies in unsupervised methods? The difference of being in an online and offline setting? How does deep learning based methods compare in complexity?

2

Methods

Inspiration:

- `http://www.nada.kth.se/~ann/exjobb/maxim_wolpher.pdf`
- [Aggarwal, 2013]

Overview

Prediction-based

- VARIMA [Aggarwal, 2013]
- Neural Networks (CNN/RNN/LSTM/GRU)
- Using PCA to get univariate forecasting [Aggarwal, 2013]

Distance and density-based

- K-Nearest Neighbor (KNN) [Li et al., 2019]
- Clustering Based Local Outlier Factor (CBLOF) [Li et al., 2019]
- Dynamic Time Warping [Aggarwal, 2013]
- Mahalanobis distance [Aggarwal, 2013]
- Isolation Forest

Probabilistic

- Hidden Markov Model [Aggarwal, 2013]
- Dynamic Bayesian Networks

Linear Model

- Principal Component Analysis (PCA) [Li et al., 2019]
- Partial Least Squares (PLS) [Li et al., 2019]
- Matrix Factorization [Aggarwal, 2013]
- Support Vector Machines [Aggarwal, 2013]

Deep Learning

- Auto Encoder (AE) [Li et al., 2019]
- Deep Autoencoding Gaussian Mixture Model (DAGMM) [Li et al., 2019]
- LSTM Encoder-Decoder (LSTM-ED) [Li et al., 2019]
- Generative Adversarial Networks (GAN) [Li et al., 2019]

Transformation to other representations

- Leveraging Trajectory Representations of Time Series [Aggarwal, 2013]
- Numeric Multidimensional Transformations [Aggarwal, 2013]
- Discrete Sequence Transformations [Aggarwal, 2013]

2.1 Prediction-based

Supervised regression techniques for predicting time-series and then finding the contextual anomalies as deviations from the predicted value is a common method. Traditional methods such as ARIMA and VARIMA can be used for the forecasting, but recently deep learning methods for predicting the values have been explored. In the multivariate case PCA is often used to reduce the problem to a univariate problem.

VARIMA

The following is cut from chapter 9.2 in [Aggarwal, 2013].

"By predicting the next value in the series contextual anomalies can be found by comparing the predicted value with the measured. This can be seen as a supervised method for unsupervised anomaly detection. The basic idea in multivariate auto-regressive models is to predict the values at each time-stamp with the past window of length p. The main difference from uni-variate regressive models is that the value at each time-stamp (for any particular series) is predicted as a linear function of all the d · p values in all the streams in the previous window of length p.

One problem with the approach is that of increased computational complexity because of the inversion of a matrix. How can one use the basic principle of multivariate regression for forecasting, while keeping the complexity to a manageable level? Two such methods have been proposed in the literature:

1. One can select a subset of streams in order to perform the regression with respect to a smaller set of variables. (9.2.2.2)

2. A fundamentally different approach is to use the notion of hidden variables to decompose the multivariate forecasting problem into a (more easily solvable) set of uni-variate forecasting problems. (9.2.2.3)"

Neural networks

Neural networks have been effective at predicting time-series. Especially recurrent neural networks. Information covering this:

- LSTM-based Encoder-Decoder for Multi-Sensor Anomaly Detection [Malhotra et al., 2016].

- Multivariate Time Series Forecasting with LSTMs in Keras `https://machinelearningmastery.com/multivariate-time-series-forecasti`

- DeepAnT: A Deep Learning Approach for Unsupervised Anomaly Detection in Time Series [**Munir2019DeepAnT:Series**]

PCA-based methods

These sources present methods and ideas of PCA-based methods for anomaly detection:

- Streaming Pattern Discovery in Multiple Time-Series [**Papadimitriou2005StreamingTime**

Regression modelling is more susceptible to noise and outliers. PCA-based methods are generally more robust. They are able to express a large number of correlated data streams into a small number of uncorrelated data streams. [Aggarwal, 2013]. Any forecasting model can be applied to the reduced hidden variables. This reduces the time and space complexity by orders of magnitude, because typical forecasting methods are quadratic or worse on the number of variables. Space complexity for multivariate AR is $\mathcal{O}(n^3 l^2)$, where l is the auto-regression window length. For AR per stream (ignoring correlations), it is $\mathcal{O}(nl^2)$. However, for SPIRIT, we need $\mathcal{O}(kn)$ space, with one AR model per stream, the total space complexity is $\mathcal{O}(kn + kl^2)$. [**Papadimitriou2005StreamingTime-series**]

2.2 Deep learning

Bibliography

Aggarwal, C. C. (2013). *Outlier analysis (book)*, pp. 1–446. ISBN: 9781461463962. DOI: `10.1007/978-1-4614-6396-2`. arXiv: `arXiv:1011.1669v3`.

Li, D., D. Chen, L. Shi, B. Jin, J. Goh, and S.-K. Ng (2019). “MAD-GAN: Multivariate Anomaly Detection for Time Series Data with Generative Adversarial Networks”. arXiv: `1901.04997`. URL: `http://arxiv.org/abs/1901.04997`.

Malhotra, P., A. Ramakrishnan, G. Anand, L. Vig, P. Agarwal, and G. Shroff (2016). “LSTM-based Encoder-Decoder for Multi-sensor Anomaly Detection”. arXiv: `1607.00148`. URL: `http://arxiv.org/abs/1607.00148`.