école normale supérieure paris—saclay

Semi-Supervised SVM for Anomaly Detection in Datasets of Time Series

Mini-Project : Machine Learning for Time Series

Steven Zheng and Lucas Haubert

Master MVA

ENS Paris-Saclay

February 22, 2024

Table of Contents

Introduction - Context

- A critical task in various domains involving time series analysis is performing an anomaly detection step, in order to optimize some other tasks relative to the domain of application.

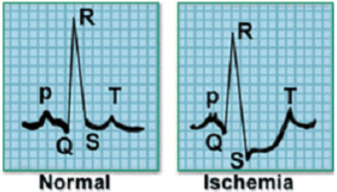

Figure: Normal heartbeat vs myocardial infarction, from [1]

- The Anomaly Detection (AD) task can be optimized with the introduction of some expert feedback.
- In this context, article [2] *How to introduce expert feedback in one-class support vector machines for anomaly detection ?* provides a semi-supervised solution in order to perform relevant AD.

Introduction - Challenges

- The paper we are working on performs anomaly detection on point data (i.e. $x \in \mathbb{R}^d$). However, time series and point data do not have the same nature.

TIME SERIES $\neq$ DATA POINTS

- Studying a time series means taking into account the fact that the values in this time series are sequenced, and that their order is very important in understanding this type of data.
- The challenge : Adapting the ν-SSVM approach to the study of time series.
- The purpose of the mini-project is to solve such a challenge, then evaluate the underlaying method with in a comparative way (against supervised, unsupervised and semi-supervised approaches).

Method - SVM

- **Purpose**: SVM is a supervised learning model used for classification.
- **Key Concept**: Identifies the best hyperplane that separates data points of different classes in a feature space.
- **Optimization Problem**:

$$\arg\min_{w,b} \left(\frac{1}{2} \|w\|^2 \right)$$

  Subject to $y_i \left(w \cdot \Phi(x_i) + b \right) \geq 1$ for each i.

- **Kernel Trick**: Used to handle non-linearly separable data by mapping input space to a higher-dimensional space.
- **Types of Kernels**:
  - Linear: $k(x_i, x_j) = x_i \cdot x_j$
  - Polynomial: $k(x_i, x_j) = (\gamma x_i \cdot x_j + r)^d$
  - Radial Basis Function (RBF): $k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$

école normale supérieure paris–saclay

Method - ν-SVM and OCSVM

- **Objective**: Enhance Anomaly Detection (AD) in datasets with both labeled and unlabeled data using semi-supervised learning.
- **Labeled Data Processing**:
  - Utilizes ν-SVM for labeled data ($\{(x_i, y_i)\}_{i=1,\ldots,n}$).
  - Incorporates a margin ρ and slack variables ξ_i to manage anomalies.
- **Unlabeled Data Processing**:
  - Employs One-Class SVM (OCSVM) for unlabeled data.
  - Finds a separating hyperplane with allowance for outliers.

Method - ν-SSVM

- **Optimization Formula**:

$$
\begin{aligned}
\underset{w\in\mathbb{R}^q,\xi\in\mathbb{R}^n,b,\rho_1,\rho_2\in\mathbb{R}}{\arg\min} \quad & \frac{1}{2}\|w\|^2 - r\rho_1 - (n-r)(\rho_2 - b) + \frac{1}{\nu_1}\sum_{i=1}^{r}\xi_i + \frac{1}{\nu_2}\sum_{i=r+1}^{n}\xi_i \\
\text{s.t.} \quad & y_i(w^T\Phi(x_i)+b) \geq \rho_1 - \xi_i, \quad i = 1,\ldots,r, \\
& w^T\Phi(x_i)+b \geq \rho_2 - \xi_i, \quad i = r+1,\ldots,n, \\
& \rho_1 \geq 0, \rho_2 \geq 0, \xi_i \geq 0, \quad i = 1,\ldots,n
\end{aligned}
\tag{ν-SSVM}
$$

- **Key Parameters**:
  - ν_1: Trust in expert feedback.
  - ν_2: Prior knowledge of anomalies in training set.

Contribution - Pre-processing

- **Why a pre-processing step ?** : We work with real-world data = possible imperfections that have nothing to do with the actual features of the data.
- **Which pre-processing and why ?** :
  - Interpolation of missing data : No needed here, as the datasets were complete
  - Detrending : Not considered here, as trends in time series can differentiate them
  - Outlier points removal : Yes, as some single outliers can impact the ν-SSVM process
  - Denoising : Yes, as the data were not pre-processed before
  - Time-frequency feature representation : Yes, as it is naive to consider time series as a set of data points
- **Key elements** :
  - Outlier points removal : Use of a rolling median filter with window size equal to 3
  - Denoising : Use of a rolling mean filter with window size equal to 3 (not too smooth)
  - Time-frequency feature representation : Use of Daubechies wavelets of various orders (deal with non stationary processes and keep a similar framework with db1, db2, db3, etc)

Contribution - Parametrization and Evaluation

- **Hyperparameters**:
  - ν_1: Bounds for the fraction of support vectors and misclassified labeled data (0.01 to 0.05 for 1-5% label errors).
  - ν_2: Lower bound for the fraction of support vectors among unlabeled data, indicating anomaly presence.
  - γ: RBF kernel parameter, $\gamma = \frac{1}{2\sigma^2}$, influencing training example range and decision boundary.
- **Methodology**:
  - Used 50% of labels in datasets.
  - Performed grid search on parameters, focusing on balanced accuracy.
  - Balanced accuracy considers the imbalance between normal and anomalous data.
- **Kernel Exploration**:
  - Experimented with polynomial and sigmoid kernels.
  - These kernels did not yield optimal results compared to RBF.
- **Outcome**:
  - Results based on optimized parameters from grid search.
  - Demonstrates the effectiveness of parameter tuning in anomaly detection.

Data

- **Data Characteristics**:
  - Labeled data for semi-supervised learning.
  - Classified into two categories: normal and anomalous time series.
  - Diversity in data: one-dimensional and multi-dimensional, varying natures (e.g., ECG, spectrograms) and lengths.
- **Data Processing**:
  - Emphasis on denoising and outlier removal (optimal rolling window size: 3).
  - For multivariate series (e.g., Heartbeat), individual time series are processed and concatenated.

Dataset	Train Size	Test Size	Length	No Classes	No Dimensions	Type
Heartbeat	204	205	405	2	61	AUDIO
ECG200	100	100	96	2	1	ECG
Strawberry	613	370	235	2	1	SPECTRO
ToeSegmentation2	36	130	343	2	1	MOTION

Results - Evaluation approach

Evaluation Approach:

- Initial analysis with a naive approach - directly applying ν-SSVM to time series data.
- Focused on balanced accuracy metric due to class imbalance.
- Comparative analysis with different preprocessing techniques.

Results - Heartbeat

Model	Accuracy	Balanced Acc	Precision	Recall	AUC	FPR
nuSSVM	0.607843	0.684927	0.903614	0.510204	0.728249	0.140351
OCSVM	0.323529	0.423201	0.591837	0.197279	0.466404	0.350877
NuSVM	0.833333	0.793054	0.884354	0.884354	0.755221	0.298246
S3VM	0.387255	0.547977	0.843750	0.183673	0.527390	0.087719

Table: Model Performance on Heartbeat data set with db1-DWT

Results - ECG

Model	Accuracy	Balanced Acc	Precision	Recall	AUC	FPR
nuSSVM	0.785	0.734654	0.722222	0.582090	0.737515	0.112782
OCSVM	0.740	0.671193	0.659574	0.462687	0.674223	0.120301
NuSVM	0.810	0.786780	0.716418	0.716418	0.822803	0.142857
S3VM	0.715	0.578330	0.916667	0.164179	0.711929	0.007519

Table: Model Performance on ECG data set with db1-DWT

Results - Strawberry

Model	Accuracy	Balanced Acc	Precision	Recall	AUC	FPR
nuSSVM	0.581892	0.560203	0.425000	0.484330	0.627094	0.363924
OCSVM	0.440488	0.353964	0.076596	0.051282	0.252022	0.343354
NuSVM	0.659207	0.653265	0.518692	0.632479	0.732223	0.325949
S3VM	0.571719	0.455387	0.163462	0.048433	0.609098	0.137658

Table: Model Performance on Strawberry data set école normale supérieure paris-saclay

Results - ToeSegmentation

Model	Accuracy	Balanced Acc	Precision	Recall	AUC	FPR
nuSSVM	0.728916	0.637481	0.463415	0.452381	0.689132	0.177419
OCSVM	0.656627	0.541859	0.317073	0.309524	0.559908	0.225806
NuSVM	0.849398	0.812596	0.688889	0.738095	0.819892	0.112903
S3VM	0.728916	0.574501	0.440000	0.261905	0.648041	0.112903

Table: Model Performance on Toe data set with denoising + outlier removal

Results - Key findings

Key Findings:

- *Heartbeat Dataset (db1-DWT)*: Balanced accuracy of **0.684927**.
- *ECG Dataset (db1-DWT)*: Improved performance, balanced accuracy of **0.734654**.
- *Strawberry Dataset*: Lower performance, balanced accuracy of **0.560203**.
- *Toe Dataset (Denoising + Outlier Removal)*: Balanced accuracy of **0.637481**.

Analysis:

- nuSSVM adaptable in semi-supervised settings.
- Demonstrated the importance of preprocessing, especially wavelet transforms, for handling time series data.
- Challenges in applying SVM to certain types of time series data, like the Strawberry dataset.

Thank You

for Listening.

Reference

Time series classification.
https://www.timeseriesclassification.com/dataset.php.

Cédric Baudoin Marc Spigai Julien Lesouple and Jean-Yves Tourneret.
How to introduce expert feedback in one-class support vector machines for anomaly detection ?, in
Signal Processing, pp. 108197.
2021.