

CLASSIFICATION OF REGIONAL IONOSPHERIC DISTURBANCE BASED ON MACHINE LEARNING TECHNIQUES

Merve Begum Terzi, Orhan Arikan, Secil Karatay, Feza Arikan, Tamara Gulyaeva

Bilkent University, Department Of Electrical and Electronics Engineering, Bilkent, Ankara, Turkey

karatays@gmail.com



INTRODUCTION

- Solar, geomagnetic, gravitational and seismic activities cause variations in the electron distribution of the atmosphere.
- The number of electrons within a vertical column of $1 m^2$ cross section, which is called as Total Electron Content (TEC) can be measured accurately by using the phase difference between transmitted satellite positioning signals such as in the Global Positioning System (GPS) [2].
- This study is concerned with investigating TEC to detect seismo-ionospheric anomalous variations induced by earthquakes.
- TEC estimated from GPS receivers is used to classify the regional and local variability that differs from global activity along with solar and geomagnetic indices.
- For the automated classification of the regional disturbances, a classification technique based on a robust machine learning technique that have found wide spread use, Support Vector Machine (SVM) is used.
- Performance of the developed classification technique is demonstrated for midlatitude ionosphere over Anatolia using TEC estimates generated from the GPS data for solar maximum year of 2011.

SUPPORT VECTOR MACHINES

- Support vector machines (SVMs) are supervised learning models used for classification in machine learning, with associated learning algorithms that analyze data and recognize patterns.
- The aim of an SVM model is to represent the samples as points in space, mapped so that the samples of the separate classes are distinguished by an apparent space that is as large as possible [1].
- After the mapping of the new samples, they are predicted to be a part of a class according to the side of the space that they correspond.

LINEAR SVM

- Given a set of n points which is the training data of the form

$$D = \{(\mathbf{x}_m, y_m) \mid \mathbf{x}_m \in R^p, y_m \in \{-1, 1\}\}_{m=1}^n$$

where each \mathbf{x}_m is a p -dimensional feature vector and y_m shows to which class the point \mathbf{x}_m belongs with a value either 1 or -1.

- The aim is to determine maximum-margin hyperplane which separates points having $y_m = 1$ from those having $y_m = -1$.

- Every hyperplane can be identified by using the set of points \mathbf{x} which satisfies:

$$\mathbf{w}^T \mathbf{x} = b$$

where \mathbf{w} states the vector which is normal to the hyperplane.

- For the cases where training data is linearly separable, two hyperplanes which separate the data can be chosen as shown below and the margin between these two hyperplanes can be maximized [2].

$$\mathbf{w}^T \mathbf{x} - b = 1 \quad \mathbf{w}^T \mathbf{x} - b = -1$$

- Since the interval between these two hyperplanes is $\frac{2}{\|\mathbf{w}\|_2}$, the aim is to minimize $\|\mathbf{w}\|_2$.

- The optimization problem becomes:

minimize $\|\mathbf{w}\|_2$ subject to:

$$y_m(\mathbf{w}^T \mathbf{x} - b) \geq 1$$

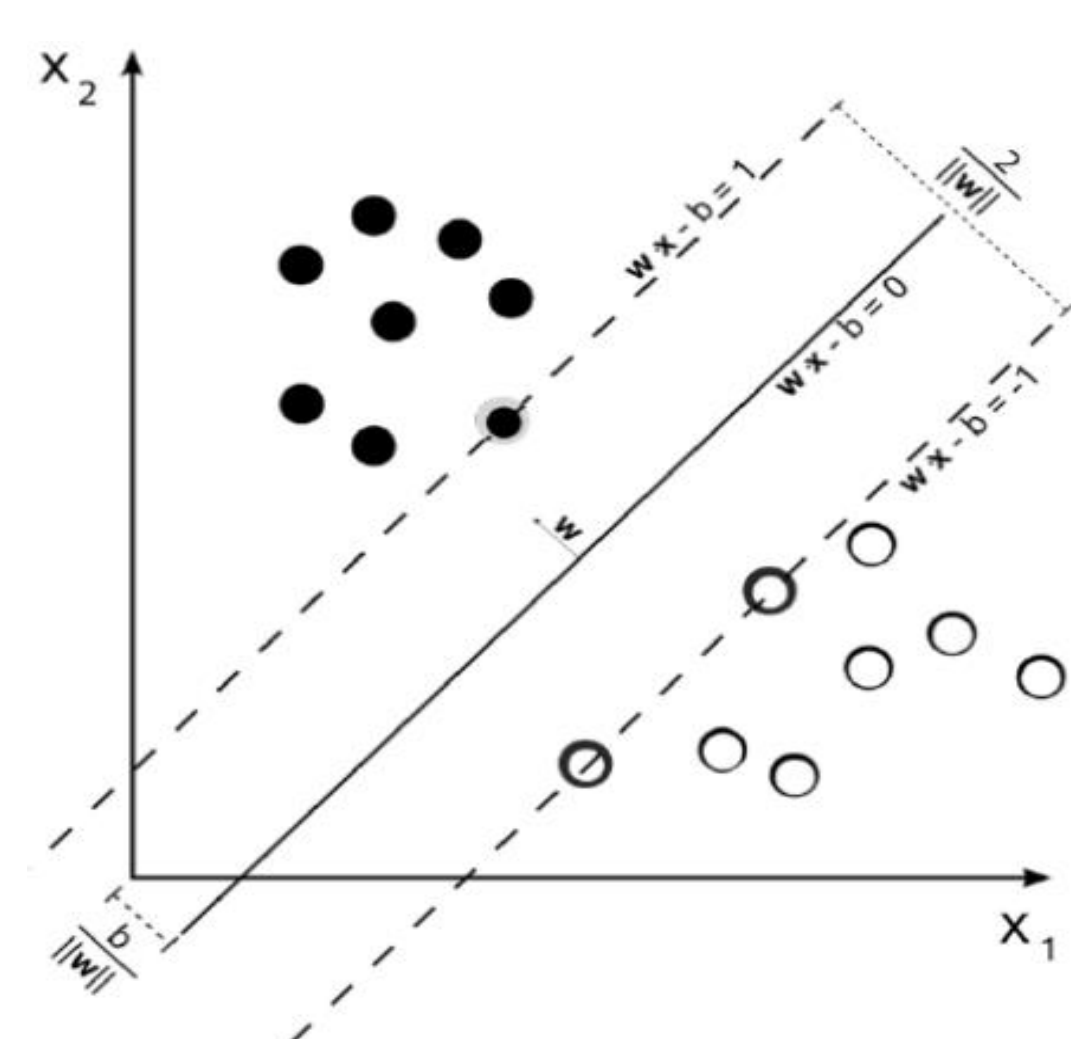


Figure 1: Margins and maximum-margin hyperplane for an SVM trained with samples from two classes. Support vectors are the samples on the margin.

- Non-negative slack variable, ε_m , that measures the misclassification level of the data \mathbf{x}_m is presented by soft margin method [3].

$$y_m(\mathbf{w}^T \mathbf{x} - b) \geq 1 - \varepsilon_m, \quad 1 \leq m \leq n$$

- Objective function is then increased by a function which penalizes non-zero ε_m , and the optimization becomes a tradeoff between a small error penalty and a large margin. The optimization problem in case of a linear penalty function is:

$$\arg \min_{\mathbf{w}, \varepsilon, b} \left\{ \frac{1}{2} \|\mathbf{w}\|_2^2 + C \sum_{m=1}^n \varepsilon_m \right\}$$

subject to

$$y_m(\mathbf{w}^T \mathbf{x} - b) \geq 1 - \varepsilon_m, \quad \varepsilon_m \geq 0$$

- By using Lagrange multipliers, constraint shown in the equation above together with the objective of minimizing $\|\mathbf{w}\|_2$ can be done by solving the following problem [4].

$$\arg \min_{\mathbf{w}, \varepsilon, b} \max_{\alpha, \beta} \left\{ \frac{1}{2} \|\mathbf{w}\|_2^2 + C \sum_{m=1}^n \varepsilon_m - \sum_{m=1}^n \alpha_m [y_m(\mathbf{w}^T \mathbf{x} - b) - 1 + \varepsilon_m] - \sum_{m=1}^n \beta_m \varepsilon_m \right\}$$

$$\beta_m, \alpha_m \geq 0$$

METHODOLOGY

- TEC variations have been analyzed using Global Ionospheric Map (GIM) data provided by the NASA Jet Propulsion Laboratory (JPL).
- GIM is constructed from a $5^\circ \times 2.5^\circ$ (longitude, latitude) grid with a time resolution of 2 hours.
- To discriminate the seismo-ionospheric perturbations from geomagnetic disturbances, the geomagnetic and solar indices (Dst and Kp indices) have been used.
- In cases where the data is linearly separable, SVMs operating with linear kernel functions are used to map the training data into kernel space.
- For the data which are not linearly separable, SVMs operating with radial basis function (RBF) kernel are employed.

RESULTS

- To obtain satisfactory predictive accuracy, the parameters of linear and RBF kernels are tuned by performing 10-fold cross-validation.
- In each cross validation fold, statistical measures of the performance of SVM classifier, such as detection rate, false alarm rate, specificity, accuracy, positive predictive value and negative predictive value are calculated.
- Performance results from the folds are then averaged to produce a single estimation which enables to choose the classifier that gives the best performance results.

Table 1: SVM Classifier Performance Results (%) for Optimum Linear Kernel Parameters and Joint Features Total Electron Content (TEC) And Kp Index.

SVM Classifier Performance Results (%)								
Earthquake	E1	E2	E3	E4	E5	E6	E7	E8
Detection Rate (P_D)	100.00	92.15	93.25	90.43	98.56	88.52	77.64	85.27
False Alarm Rate (P_{FA})	0	5.36	4.74	7.21	0	5.36	11.02	8.48
Specificity	100.00	87.42	93.25	92.15	100.00	86.27	79.45	84.28
Accuracy	100.00	93.26	91.57	89.37	95.39	89.62	76.39	86.39
Positive Predictive Value	100.00	86.37	93.52	81.63	98.12	79.42	73.25	76.28
Negative Predictive Value	100.00	98.24	95.26	91.46	100.00	86.26	74.29	82.27
AUC (ROC)	100.00	99.36	98.32	93.76	98.23	90.43	80.12	87.31
AUC (PR)	100.00	96.43	97.42	94.07	97.35	89.20	77.94	85.62
Average Cross-Validation Error	0	4.68	3.92	4.85	1.35	7.36	11.38	9.47

Table 2: SVM Classifier Performance Results (%) For Optimum Linear Kernel Parameters And Joint Features Total Electron Content (TEC) And Dst Index.

SVM Classifier Performance Results (%)								
Earthquake	E1	E2	E3	E4	E5	E6	E7	E8
Detection Rate (P_D)	100.00	94.56	95.48	92.41	100.00	91.54	81.17	84.32
False Alarm Rate (P_{FA})	0	2.04	2.47	5.37	0	3.52	10.13	9.46
Specificity	100.00	90.76	96.41	95.36	100.00	88.63	82.87	85.23
Accuracy	100.00	92.75	94.12	91.10	97.13	90.48	80.31	85.35
Positive Predictive Value	100.00	89.42	92.34	84.58	98.76	82.65	76.18	80.14
Negative Predictive Value	100.00	100.00	95.03	90.74	100.00	87.48	79.71	84.46
AUC (ROC)	100.00	100.00	100.00	95.47	100.00	91.35	83.16	89.32
AUC (PR)	100.00	96.26	97.42	94.56	98.83	92.63	81.74	87.34
Average Cross-Validation Error	0	3.12	3.25	4.37	0.61	5.74	10.48	8.13

(See QR code link for earthquake information table)

Figures 2 and 3 shows, SVM classification results for Tabanlı, Van earthquake (E1) by using joint features TEC data and Dst or Kp indices, respectively.

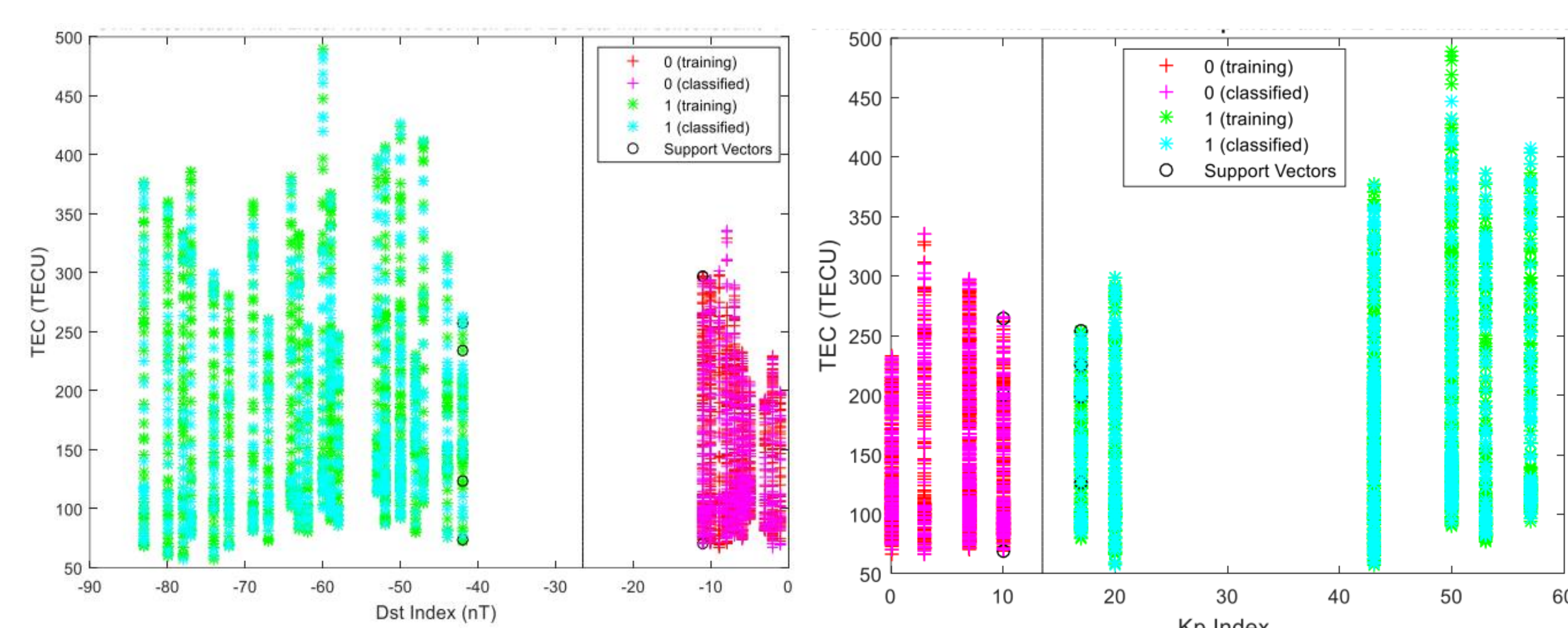


Figure 2: SVM classification with linear kernel, optimum kernel parameters and joint features Total Electron Content and Dst Index.

Figure 3: SVM classification with linear kernel, optimum kernel parameters and joint features Total Electron Content and Kp Index.

CONCLUSION

- In this study, SVMs are used for the automated classification of the regional disturbances.
- Performance of the developed classification technique is demonstrated for midlatitude ionosphere over Anatolia using TEC estimates generated from the GPS data for solar maximum year of 2011.
- To discriminate the seismo-ionospheric perturbations from geomagnetic disturbances, the geomagnetic and solar indices (Dst and Kp indices) have been used.
- As a result of implementing the developed classification technique to the GIM TEC data, it is shown that SVM can be a suitable learning method to detect the anomalies in Total Electron Content (TEC) variations.

ACKNOWLEDGEMENT

This study is supported by TUBITAK 114E541, 115E915 and and joint TUBITAK 114E092 and AS CR 14/001 projects. The GIM-TEC data is obtained from IGS Iono Working Group Data Analysis Center of Jet Propulsion Laboratory at <http://cddis.gsfc.nasa.gov/gps/products/ionex/>. The Kp and Dst indices are obtained from Data Analysis Center for Geomagnetism and Space Magnetism, Kyoto University, Japan (http://wdc.kugi.kyoto-u.ac.jp/dst_realtime/index.html and <http://wdc.kugi.kyoto-u.ac.jp/kp/index.html>).

REFERENCES

- Nayir, H., F. Arikan, O. Arikan, and C. B. Erol, Total Electron Content estimation with Reg-Est, J. Geophys. Res., 112, 2007.
- Vapnik, V.N., 2000, The Nature of Statistical Learning Theory, 2. Baski, Springer-Verlag, New York.
- Liu, J. Y., Chuo, Y. J., Shan, S. J., Tsai, Y. B., Chen, Y. L., Pulnits, S. A., and Yu, S. B.: Preearthquake ionospheric anomalies registered by continuous GPS TEC measurements, Ann. Geophys., 22, 1585–1593, doi:10.5194/angeo-22-1585-2004, 2004.
- Thissen, U., Brakel, R. V., deWeijer, A. P., Melssen, W. J., and Buydens, L. M. C.: Using support vector machines for time series prediction, Chemometrics and Intelligent Laboratory Systems, 69, 35–49, 2003.

