Analysis and Prediction of Ionospheric Scintillation through Data Mining Techniques

Luiz Felipe Campos de Rezende, INPE - Brazil
Stephan Stephany, INPE - Brazil
Eurico Rodrigues de Paula, INPE - Brazil
Ivan Jelinek Kantor, INPE - Brazil
Márcio T.A.H. Muella, INPE - Brazil

This paper was prepared for presentation during the 11th International Congress of the Brazilian Geophysical Society held in Salvador, Brazil, August 24-28, 2009.

Abstract

This paper describes the application of Data Mining to the analysis and prediction of ionospheric scintillation over the Brazilian territory. Ionospheric scintillation is a phenomenon that occurs in the equatorial region. It affects the telecommunications and Global Positioning System (GPS) accuracy for positioning and aerial navigation. Data Mining can be defined as the process of extracting useful information or finding hidden patterns from databases using techniques of statistics, artificial intelligence and patterns recognition.

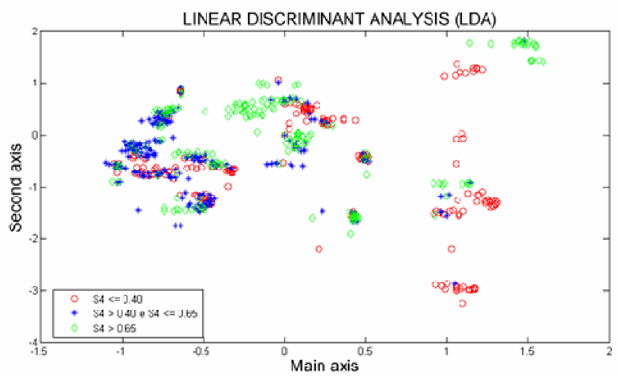
Introduction

In the post-sunset equatorial ionosphere, plasma depleted regions/bubbles with associated irregularity structures of scale sizes varying from centimeters to kilometers are generated due to plasma instability processes [de Paula et al., 2007]. The ionospheric irregularities present a large dependence of the solar flux, the local time, the season, the latitude and longitude and the magnetic disturbances [de Paula et al., 2007]. Ionosphere scintillation occurs when a radio wave crosses the ionosphere and suffers a distortion of phase and amplitude. This produces amplitude and phase fluctuations. Several studies have demonstrated that the equatorial ionospheric scintillations affect the performance of GPS receivers [Kintner et al., 2001]; [Bandyopadhayay et al., 1997]. The aim of this work is to propose a method for 1-hour prediction of ionospheric scintillation using Data Mining. Such prediction would be useful for issuing aerial navigation warnings about the degradation of GPS due to scintillation. GPS services require availability, accuracy, continuity and integrity.

Method

Data Mining can be defined as the process of extracting useful information or finding hidden patterns from databases using techniques of statistics, artificial intelligence and patterns recognition. Useful patterns allows to make non-trivial predictions on new data [Witten et al., 2000]. In order to find meaningful patterns, a set of parameters was selected for this work. Such parameters are the scintillation index in the magnetic equator, the scintillation index in São José dos Campos with a 1-hour delay, the magnetic activity given by the Kp index, the solar flux (F10.7), and the vertical drift velocity of plasma in the equator. The scintillation intensity is quantified by the S4 index that is defined as the normalized variation of the signal intensity of GPS satellites [Beach, 1998];[de Rezende et al., 2007]. The receiver used to get ionospheric scintillation data was the GEC-Plessey GPS card with specific software developed by Cornell University to acquire data at a 50 Hz sampling rate and single frequency L1 (1,575.42 MHz). The Kp index and solar flux are obtained from the website of the University of Kyoto (http://swdcwww.kugi.kyoto-u.ac.jp/kp/index.html). The vertical drift velocity is given by the DGS256 Digisonde - Digital Portable Sound (DPS) installed in São Luís (Northeast of Brazil). The objective is to get a 1-hour prediction of the ionospheric scintillation in São José dos Campos. This city is in the Southeast of Brazil, a region where scintillation is very severe due to the proximity to the ionospheric Anomaly Peak and affects drastically GPS receivers signal. The data used in this work corresponds to the solar maximum period 2001-2002 from November to March. The sun has a solar cycle of approximately 11 years during which there are variations of the sunspots and solar flux.

The prediction of the ionospheric scintillation is related to a nonlinear model and non-separable variables as shown in Figure 1 by the linear discriminant analysis (LDA) algorithm [Duda et al., 2001] implemented in MatLab. Some predictors were then used in this work to perform the inference of the scintillation in São José dos Campos with a 1-hour delay. These predictors are a Multilayer Perceptron neural network, Bagging and K-star algorithms.



Bagging (bootstrap aggregating) algorithms generate multiple versions of a predictor and using these to get an aggregated predictor [Breiman, 1994]. New training sets are generated by repeatedly selecting uniformly at random and with replacement of these examples from the original training set. The aggregation scheme averages over the versions when predicting a numerical outcome. The decision tree used in the bagging was the REPtree that builds a decision/regression tree using information gain/variance.

The KStar algorithm is an instance-based learner that uses entropy as a distance measure between two instances. This was motivated by the information theory and such distance is based on the Kolmogorov complexity. The intuition is that the distance between instances be defined as the complexity of transforming one instance into another [Cleary et al., 1995]. It was developed in the scope of the Weka project of the University of Waikato (http://www.cs.waikato.ac.nz/ml/weka/).

Preliminary results

The algorithms mentioned in the previous section were evaluated separately. In a first step, each algorithm was trained with a training dataset and further tested using a different test dataset. Both datasets include data acquired every 5 minutes for daily period of 4 hours during which there is scintillation. The S4 data at the Equator and at the Anomaly peak is very noisy and demanded a standard smoothing technique. The training dataset has 8354 instances and the test dataset only 32 randomly-chosen instances. In the training phase, a 10-fold cross validation is performed and results are shown in Table 1. The MultilayerPerceptron (MLP) neural network showed a poor performance in this phase, while results for KStar and Bagging were encouraging. In the test phase, both the KStar and Bagging algorithms were able to estimate the curve for the S4 index at the anomaly peak. This curve was generated by shuffling 32 instances of the full dataset. The estimations are shown in Figures 2 and 3 for the Baggin and KStar algorithms, respectively.

Table 1 Results for 10-fold cross validation training

Algorithm	Correlation coefficient	Relative absolute error (%)	Root relative squared error (%)
Bagging	0.9974	4.9695	7.296
KStar	0.9973	5.2158	7.6773
MLP	0.6883	70.7241	74.6547

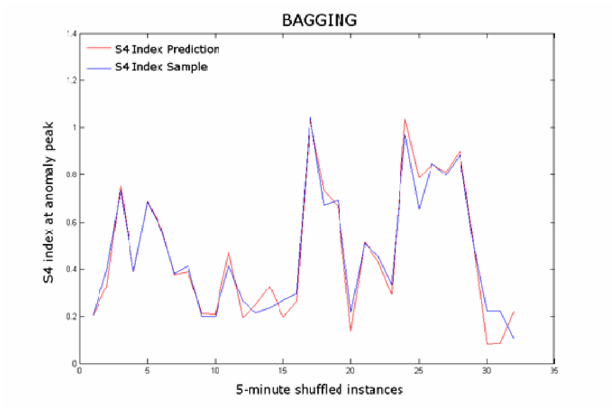


Figure 2: A prediction with Bagging

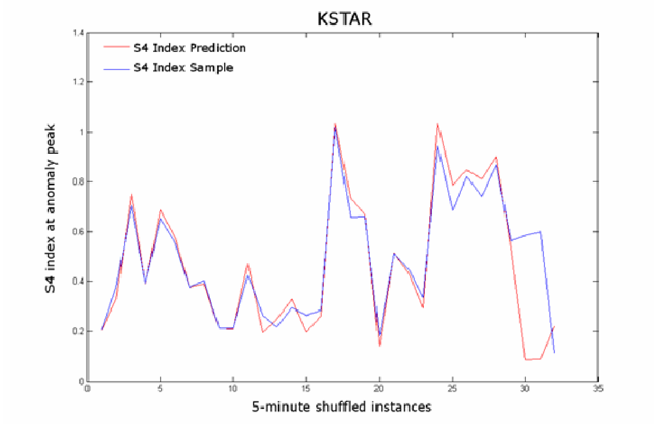


Figure 3: A prediction with KStar

Conclusions and remarks

Preliminary results shows that the proposed methodology is feasible with both the Bagging and KStar algorithms. The S4 index at the Anomaly Peak was estimated for a shuffled set of 32 instances. It is important to note that S4 data in this work is used with resolution of 5 minutes, but other parameters have a single daily value that precluded a good estimation (not shown here) in the case of a curve composed by instances that are sorted in a time sequence, i.e. forming the daily sequence of S4 values. In order to perform such estimation, we are considering the use of 1st and 2nd derivatives of the S4 curve.

Acknowledgments

The authors acknowledge Dra. Inez Staciarini Batista and Maria Goreti S. Aquino for providing data from DGS256 Digisonde. Luiz Felipe Campos de Rezende is grateful to FAPESP for support to this work, under Process 2008/03675-0.

References

Bandyopadhayay, T., Guha, A., Dasgupta, A., Banerjee, P., Bose, A., “Degradation of navigational accuracy with global positioning system during periods of scintillation at equatorial latitudes”, Electronics Letters, June, 5, Vol. 33, num. 12, 1997.

Beach, T.L., Global positioning studies of equatorial scintillations, PhD Dissertation, Cornell University, Ithaca, NY, USA, 1998.

Breiman, L., Technical Report num. 421 - Department of Statistics – University of California, 1994.

Cleary, J.G., Trigg, L.E., K*: An Instance-based Learner Using an Entropic Distance Measure. In Proceedings of the 12th International Conference on Machine Learning, California, USA, p. 108-114, 1995.

de Paula, E. R., Kherani, A.E., Abdu, M. A., Batista, I. S., Sobral, J. H. A., Kantor, I.J., Takahashi, H., de Rezende, L.F.C., Muella, M.T.A.H., Rodrigues, F.S., Kintner, P.M., Ledvina, B.M., Mitchell, C., and Groves, K.M., Characteristics of the ionospheric irregularities over

Brazilian longitudinal sector, Indian Journal of Radio & Space Physics, New Delhi, Vol. 36, p. 268-277, 2007.

Duda, R.O., Hart, P.E., Stork, D.G., Pattern Classification, John Wiley & Sons, United States, 2001.

Kintner, P.M., Kil, H., Beach, T.L., de Paula, E.R., Fading timescales associated with GPS signal and potential consequences, Radio Science, 36 (4), p731-743, 2001.

Rezende, L.F.C., de Paula, E.R., Batista, Kantor, I. J., Kintner, P. M., Mapping and Survey of Plasma Bubbles over Brazilian Territory, The Journal OF Navigation, 60, 69–81, 2007.

Witten, I.H., Frank, E., Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann Publishers, São Francisco, United States, 2000.