Neural Network-based Prediction of Solar Activities

Rami S. R. QAHWAJI
EIMC, University of Bradford
Bradford, BD71DP, U.K.

and

Tufan COLAK
EIMC, University of Bradford
Bradford, BD71DP, U.K.

ABSTRACT

A data mining system designed to associate previous solar flares with sunspot groups using databases from Nobeyama Radioheliograph and the National Geophysical Data Center, and associated data used for training verification, and comparison of several Neural Networks topologies which can be used with an automated solar activity prediction system in the future. The data mining system manages to associate 272 flare and sunspot groups out of 1628 flares and 65363 sunspot groups using the degree of correspondence between their locations and time data. The cascade feed forward backpropagation trained network provided the optimum performance and 85% correct prediction for the possible occurrence of a solar flare was obtained. In addition, 78% of the class of the occurring flares are predicted correctly.

Keywords: Neural Networks, Solar Flares, Sunspots

1. INTRODUCTION

The term "space weather" refers to adverse conditions on the Sun that may affect space-borne or ground-based technological systems and can endanger human health or life [1].The importance of space weather is increasing day after day because of the way solar activities affect life on Earth and it will continue to increase as we rely more and more on different communication and power systems. The most dramatic solar events affecting the terrestrial environment are solar flares and Coronal Mass Ejections (CMEs) [2]. These solar eruptions can spew vast quantities of radiation and charged particles into space [3] that could cause serious damage to aircraft electronics and satellites, telecommunication and radio navigation systems, pipelines and electric power distribution grids. These solar eruptions should be predicted if an early warning system is to be designed. A long standing problem in this field is the exact degree of correlation between the solar features and solar activities. In this paper, we aim to approach this problem from the machine learning point of view. The correlation between sunspots and solar flares is investigated using neural networks.

Machine learning and data mining are not widely applied to solar data. A recent survey describing the imaging techniques used for solar feature recognition is introduced in [4]. [5] presented a method for the automatic detection of solar flares using the multi-layer perceptron (MLP) with back-propagation training rule. A supervised learning technique that required a large number of iterations was used. In [6], a comparison based on the classification performance for features extracted from solar flares is carried out between Radial Basis Function (RBF), Support Vector Machine (SVM) and MLP methods. Each flare is represented using nine features. However, these features provide no information about the position, size and verification of solar flares. Neural Networks (NNs) were used in [7] for filament recognition in solar images and in [8] neural networks (NNs) were used after image segmentation to verify the regions of interest, which were solar filaments.

This paper is organised as follows: the solar data used in this paper is described in Section 2. The NN topology and design are discussed in Section 3. Section 4 is devoted to the practical implementation and the evaluation of the performance. Finally, the concluding remarks are given in Section 5.

2. SOLAR DATA

Two different catalogues are used for this research: the Nobeyama Radioheliograph (NR) flares event list [9] and the National Geophysical Data Center (NGDC) (sunspot group data) [10]. NR flare list includes data for M and X class flares. Flares are classified according to their x-ray brightness in the wavelength range from 1 to 8 Angstroms. *C*, *M*, and *X* class flares can effect earth. *C*-class flares are moderate flares with few noticeable consequences on Earth (i.e., minor geomagnetic storms). *M*-class flares are large; they generally cause brief radio blackouts that affect Earth's Polar Regions by causing major geomagnetic storms. *X*-class flares can trigger planet-wide radio blackouts and long-lasting radiation storms. This catalogue supplies information about dates, starting and ending times, location and x-ray classification of detected flares from NR. Not all the flares have location data, so in this study only flares with location data are included.

NGDC hold records of sunspot groups reported from several observatories from all around the world, supplying their location, time, physical properties and classification data. Two classification systems exist for sunspots: McIntosh and Mt. Wilson. McIntosh classification depends on the size, shape and spot density of sunspots, while the Mt. Wilson classification [11] is based on the distribution of magnetic polarities within spot groups [12]. The McIntosh classification is the standard for the international exchange of solar geophysical data. It is a modified version of the Zurich classification system developed by Waldmeir. The general form of the McIntosh classification is *Zpc* where, *Z* is the modified Zurich class, *p* is the type of spot, and *c* is the degree of compactness in the interior of the group. Mt. Wilson classification consists of letters taken from the Greek alphabet from alpha to delta and their different combination.

3. THE TOPOLOGY AND DESIGN OF NN

The NN has proven to be a very good tool for solving many real-life problems. The efficient implementation of the NN requires training sessions and experiments that are usually long [13]. The training process depends on the training vector and on the topology of the NN. The NN manages to converge if the training data are adequate to create the appropriate discriminations between the different output classes.

Several topologies are investigated for this study such as: Elman backpropagation, feed-forward backpropagation and cascade feed forward backpropagation. The cascade feed forward backpropagation trained network provided the best results in terms of convergence time and optimum network structure. In this network, the first layer has connecting weights with the input layer. Each subsequent layer has weights connecting it to the input layer and all previous layers. The backpropagation training algorithm is widely used to train the multi-layer perceptron (MLP) networks because it provides high degrees of robustness and generalisation [14].

4. PRACTICAL IMPLEMENTATION

The flare and sunspot group data, from 01/01/1992 to 01/01/2005 are used for training. The degree of correspondence between flares and sunspots was determined based on their locations and time data. The software manages to associate 272 flare and sunspot groups out of 1628 flares and 65363 sunspot groups. The total number of samples is 500 because the data corresponding to 228 non-flaring sunspot groups is added.

The NN training and testing was carried out based on the statistical Jack-knife technique, which is usually implemented to provide a correct statistical evaluation for the performance of the classifier, when implemented on a limited number of samples. This technique divides the total number of samples into 2 sets: a training set and a testing set. Practically, a random number generator decides which samples are used for the training of the NN and which are kept for testing it. The classification error depends mainly on the training and testing samples. For a finite number of samples, the error counting procedure can be used to estimate the performance of the classifier [15]. For our case four experiments were carried out. In each experiment, 80% of the samples were randomly selected and used for training while the remaining 20% were used for testing.

For the training stage 400 input data for the NN training is used. For each sample, the training vector consists of 10 elements and is divided into 2 parts: input and target. The input part has 8 values representing McIntosh classification (3 values) and individual values representing Mt.Wilson classification, length, area, latitude and longitude. The 3 values for McIntosh classification are modified Zurich class, type of largest spot and the sunspot distribution. The target part consists of 2 values. The first target value is used to predict whether the sunspot is going to produce a flare or not. The other value is used to determine whether it is going to produce a X class flare, M class flare or no flare.

As illustrated earlier, the cascade feed forward backpropagation trained network provides the optimum performance for our case. One hidden layer with 31 neurons was used and the NN managed to converge to the normalized system error, which was 0.001. Four experiments were carried out using different training and testing data sets. Similar performances were obtained for each experiment. Overall, the average of these performances was considered. It was found that 85% correct prediction for the possible occurrence of a solar flare was obtained. In addition, the class of the flares are predicted 78% correctly.

5. CONCLUSIONS

Our findings about the association between sunspots and solar flares are in accordance with [16], [17], [18]. It also shows that an automated system can be designed with NNs to predict important flare events. The solar data used in this paper were created subjectively by solar physicists. Because of human related factors (i.e., fatigue, personal judgments), these data many not be 100% compatible. However, we believe that the results obtained here are very important because they coincide with previous research and NNs were able to establish the correlation between sunspots and solar flares. This is a first step toward constructing a fully automated system that would accept solar images, detect solar features (i.e., sunspots, filaments, etc), classify them and then predict the possible occurrence of solar activities in the short term. To ensure the success of such system, we are going to extend our experiments by using other machine learning algorithms such as support vector machine (SVM). It has become a recent trend in machine learning to compare the performance of SVM and NN, as reported in [6],[19],[20],[21], and [22]. In these papers it was reported that SVM has outperformed NNs. These applications range from recognition of sleep spindles in EEG [19], assessment of hyperspectral data [22], credit rating analysis [20] and olfactory signals recognition [21]. On the short run we aim to compare both algorithms to determine the best learning algorithm for our application.

6. REFERENCES

[1] H. Koskinen, et al., Space Weather Effects Catalogue, in ESA Space Weather Programme Feasibility Studies. 2001, FMI, QinetiQ, RAL Consortium.

[2] M. Pick, C. Lathuillere, and J. Lilensten, Ground Based Measurements, in ESA Space Weather Programme Feasibility Studies. 2001, Alcatel-LPCE Consortium.

[3] D. Lenz, "Understanding and Predicting Space Weather", **The Industrial Physicist**, 2004. 9(6): pp. 18-21.

[4] V. Zharkova, et al., "Feature Recognition in Solar Images", **Artificial Intelligence Review IAR Journal**, Accepted 2004.

[5] R.A.F. Borda, et al., "Automatic solar flare detection using neural network techniques", **Solar Physics**, 2002. 206(2): pp. 347-357.

[6] M. Qu, et al., "Automatic solar flare detection using MLP, RBF, and SVM", **Solar Physics**, 2003. 217(1): pp. 157-172.

[7] V. Zharkova and V. Schetinin, "A Neural Network Technique for Recognition of filaments in solar images", **Proceedings of the seventh International Conference on Knowledge-Based Intelligent Information & Engineering Systems KES'03**, 2003: pp. 148-154.

[8] R. Qahwaji and T. Colak, "Automatic Detection And Verification of Solar Features", **International Journal of Imaging Systems & Technology**, 2006. 15(4): pp. 199 - 210.

[9] N. RadioheliographAvailable from: http://solar.nro.nao.ac.jp/norh/html/goes_event/, Date Accessed: 2006

[10] National Geophysical Data Center. Available from: ftp://ftp.ngdc.noaa.gov/STP/SOLAR_DATA/SUNSPOT_REGIONS/USAF_MWL/ Date Accessed: 2006

[11] G.E. Hale, et al., "The Magnetic Polarity of Sun-Spots", **Astrophysical Journal**, 1919. 49: pp. 153.

[12] G.R. Greatrix and G.H. Curtis, "Magnetic Classification Of Sunspot Groups", **Observatory**, 1973. 93(994): pp. 114-116.

[13] Q. Ma, et al., "Principal Component Analysis and Artificial Neural Networks Applied to the Classification of Chinese Pottery Neolithic age", **Analytica Chimica Acta Journal**, 2000.406: pp. 247-256.

[14] J. Kim, et al., "Linear And Non-Linear Pattern Recognition Models For Classification Of Fruit From Visible-Near Infrared Spectra", **Chemometrics And Intelligent Laboratory Systems**, 2000. 51: pp.,201-216.

[15] K. Fukunaga, **Introduction to Statistical Pattern Recognition," Academic Press, New York, 1990.** 1990, New York: Academic Press.

[16] P.S. McIntosh, "The Classification Of Sunspot Groups", **Solar Physics**, 1990. 125(2): pp. 251-267.

[17] C.S. Warwick, " Sunspot Configurations and Proton Flares ", **Astrophysical Journal**, 1966. 145: pp. 215.

[18] K. Sakurai, "On the magnetic configuration of sunspot groups which produce solar proton flares", **Planetary and Space Science**, 1970. 18(1): pp. 33.

[19] N. Acir and C. Guzelis, "Automatic recognition of sleep spindles in EEG by using artificial neural networks", **Expert Systems With Applications**, 2004. 27(3): pp. 451-458.

[20] Z. Huang, et al., "Credit rating analysis with support vector machines and neural networks: a market comparative study", **Decision Support Systems**, 2004. 37(4): pp. 543-558.

[21] C. Distante, N. Ancona, and P. Siciliano, "Support vector machines for olfactory signals recognition", **Sensors And Actuators B-Chemical**, 2003. 88(1): pp. 30-39.

[22] M. Pal and P.M. Mather, "Assessment of the effectiveness of support vector machines for hyperspectral data", **Future Generation Computer Systems**, 2004. 20(7): pp. 1215-1225.