Automated Prediction of Solar Flares Using Neural Networks and Sunspots Associations

T. Colak and R. Qahwaji

Department of Electronic Imaging and Media Communications, University of Bradford
Richmond Road, Bradford BD7 1DP, England, UK.
E mail: t.colak@bradford.ac.uk; r.s.r.qahwaji@brad.ac.uk

Abstract. An automated neural network-based system for predicting solar flares from their associated sunspots and simulated solar cycle is introduced. A sunspot is the cooler region of the Sun's photosphere which, thus, appears dark on the Sun's disc, and a solar flare is sudden, short lived, burst of energy on the Sun's surface, lasting from minutes to hours. The system explores the publicly available solar catalogues from the National Geophysical Data Center to associate sunspots and flares. Size, shape and spot density of relevant sunspots are used as input values, in addition to the values found by the solar activity model introduced by Hathaway. Two outputs are provided: The first is a flare/ no flare prediction, while the second is type of the solar flare prediction (X or M type flare). Our system provides 91.7% correct prediction for the possible occurrences and, 88.3% correct prediction for the type of the solar flares.

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 established effects of space weather activities on our daily lives can be summarized as follows:

Ground based systems: Induced electric fields and currents can disrupt the normal operation of high voltage power transmission grids, pipelines, telecommunications cables, metallic oil and gas pipelines and railway signaling [2]. The great geomagnetic storm of March 13, 1989 closed down the entire Hydro Quebec system [3]. [4] predicted that a credible electric power outage could result in a direct loss to the US Gross Domestic Product of $3 - $6 billion. [5] predicted that timely forecasts could save the US power industry $365M per year.

Communications systems: Wireless communications systems suffer from interruption of service like frequency jamming and dropped communications due to radio

bursts caused by solar microwave emissions [6]. Solar activities can produce X-rays that disrupt point-to-point high frequency radio communications and radio noise that interferes with communications and radar systems [7].

Space based systems: Adverse space weather conditions can cause anomalies and system failures and increased drag on the movement of satellites and spacecraft leading to slow-downs, changes in orbits and shorter life-times of missions. Other radiation hazards include direct collision damage and/or electrical defects, caused by charged particles [8]. On 19 May 1998, the PanAmSat Corporation's Galaxy 4 satellite experienced a failure in its altitude-control system, leading to the suspension of paging service for 45 million people [9].

There have been noticeable developments recently in solar imaging and the automated detection of various solar features, by: [10], [11], [12], [13], [14] and [15].Despite the recent advances in solar imaging, machine learning and data mining have not been widely applied to solar data. [16] described a method for the automatic detection of solar flares from optical H-alpha images using the multi-layer perceptron (MLP) with back-propagation training rule. In [17], the classification performance for features extracted from solar flares is compared using Radial Basis Functions (RBF), Support Vector Machines (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 [18] for filament recognition in solar images. However, machine learning algorithms are still not applied properly for the automated prediction of solar flares and space weather activities. This is a very challenging task because of the following reasons:

- There are an increasing number of space missions and ground based observatories providing continuous observation of the Sun at many different wavelengths. We are becoming "data rich" but without automated data analysis and knowledge extraction techniques, we continue to be "knowledge poor".

- A long standing problem in solar physics is establishing a correlation between the occurrence of solar activity (e.g., solar flares and coronal mass ejections (CMEs)) and solar features (sunspots, active regions and filaments) observed in various wavelengths.

- An efficient prediction system requires the successful integration of solar physics, machine learning and maybe solar imaging.

- There is no machine learning algorithm that is known to provide the "best" learning performance especially in the solar domain. In most cases, empirical studies, in a manner similar to [17], must be carried out to compare the performances of these algorithms before the final decision on which learning algorithm to use can be made.

A first attempt in addressing these challenges is reported in the recent work of [19] where different neural network (NN) topologies were studied to determine the best NN topology to process sunspots and associate them with solar flares. The findings of this work are used in this paper to address the challenges highlighted above. In general, we aim to investigate the degree of correlation between sunspot classes and the occurrence of solar flares that can affect our life on Earth using neural networks and timing information that represent the solar activity.

This paper is organized as follows: the data used in this paper is described in Section 2. The NN topology is discussed in Section 3. Section 4 is devoted to the practi-

cal implementation and the evaluation of the performance. Finally, the concluding remarks are given in Section 5.

2 Data

2.1 Sunspots and Flare Catalogues

A sunspot is the cooler region of the Sun's photosphere which, thus, appears dark on the Sun's disc, and a solar flare is sudden, short lived, burst of energy on the Sun's surface, lasting from minutes to hours. Solar flare research has shown that flares are mostly related to sunspots and active regions [20],[21], [22]. There are many publicly available catalogues which includes the information about flares and sunspots occurred in the past.

The sunspot group catalogue and solar flare catalogue are used from the National Geophysical Data Center (NGDC) [23] for this research. Flare catalogue includes data for all type of detected 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, the National Oceanic and Atmospheric Administration (NOAA) number of the corresponding active region and x-ray classification of detected flares. NOAA numbers active regions consecutively since January 5, 1972 as they are observed on the Sun. An active region must be observed by two different observatories before it is given a number or a flare must be observed to occur in it [24]. Not all the flares have their related NOAA number, so in this study only flares with associated NOAA numbers 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 [25] is based on the distribution of magnetic polarities within spot groups [26]. 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.

2.2 Solar Cycle Prediction Data

Besides the flare and sunspot group associations, 'monthly average sunspot number' generated using Equation 1 from [27] as a part of the NN inputs. The comparison of real and generated 'monthly average sunspot number' for a given date is given in Fig. 1. The values generated by Equation 1 are more suitable for our research instead of historical data because with this Equation we can generate predicted 'monthly average sunspot numbers' and use it in our trained NN system. In Equation 1, parameter a represents the amplitude and is related to the rise of the cycle minimum, b is related to the time in months from minimum to maximum; c gives asymmetry of the cycle; and to denotes the starting time. More info about Equation 1 can be found in [27].

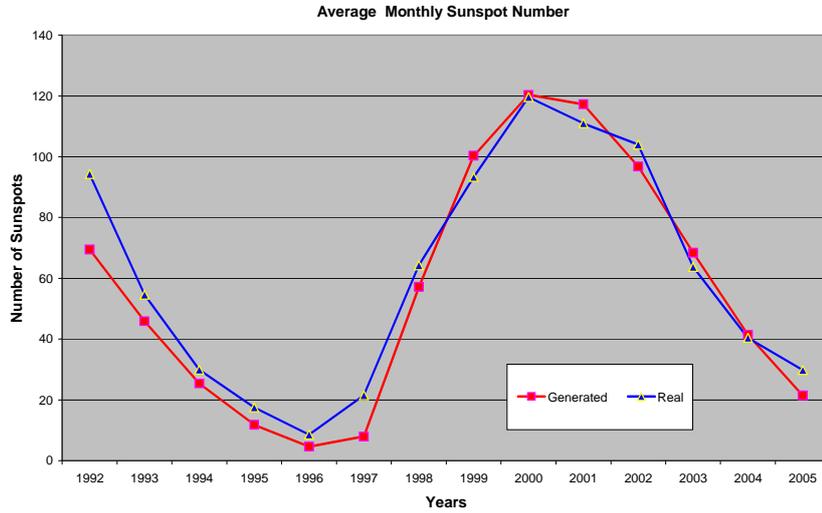$$f(t) = a(t - t_0)^3 / \{\exp[(t - t_0)^2 / b^2] - c\} \tag{1}$$

Fig. 1. The comparison of real and generated 'monthly average sunspot number' between 1992 and 2005.

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 that 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. The topology of the NN also plays an important role in the training process. If the network topology is too small then most properly the network will not be able to converge. On the contrary, if the network is too large and the training examples are presented many times, the network focuses on the singular statistical

characteristics of the training set and looses its generalization ability. A network should be large enough to learn and small enough to generalize [28], [29]. The training is improved if the NN is optimized. A NN is optimized if the best hidden layer nodes and the optimum learning time are reached [30]. In general, two hidden layers are sufficient to perform any classification task although real-life problems are often much simpler and can be solved with a single hidden layer [31]. However, once trained the NN provides fast response.

In [19], several NN topologies were evaluated and it was found that the cascade forward NN with backpropagation training provides the best performance in terms of convergence time, optimum network structure and recognition performance. In cascade forward NN, the first layer has weights coming from the input and each subsequent layer has weights coming from the input and all previous layers.

4 Practical Implementation

4.1 The Training Process

The associated flare and sunspot group data, from 01/01/1992 to 31/12/2005 are used for training. The degree of correspondence between flares and sunspots was determined based on their NOAA region number and time. The software we created manages to associate 1425 M and X flares and sunspot groups out of 29343 flares and 110241 sunspot groups. The total number of samples is 2882, where 1425 samples represent sunspots that produced flares and the remaining samples represent the distinct sunspots that existed in non-flaring days and not related to any sunspot groups within the previous flaring sunspot samples.

For each sample, the training vector consists of 6 elements and is divided into 2 parts: input and target. The input part has 4 values representing McIntosh classification (3 values) of sunspots and the simulated number of monthly average sunspots number generated based on the Hathaway's model [27]. 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 target value is used to determine whether the predicted flare is an X or M class flare.

4.2 Evaluating the performance

The NN training and testing was carried out based on the statistical Jackknife 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

[32]. In each experiment, 80% of the samples were randomly selected and used for training while the remaining 20% were used for testing. Hence, the number of samples used in the training is 2305, while 577 samples are used in the testing of the NN.

As illustrated in [19], the cascade forward backpropagation trained NN provides the optimum performance for our case. However, before applying Jackknife technique, the number of hidden nodes was found empirically. We have started with 1 hidden node and continuously were increasing the number of hidden nodes until 35 hidden nodes were reached. Every time a new number of hidden nodes were used, the error rate and the recognition rate were recorded. After carrying out all the empirical experiments it was found that optimum performance was reached with 9 hidden nodes.

After the hidden nodes number was set to 9, ten training and testing experiments, based on the Jackknife technique were carried out to evaluate the NN performance, as illustrated in Table 1. On average, our system can provide 91.7% correct prediction for the possible occurrence of a solar flare and it can predict the class of this flare correctly in 88.3% of all cases.

Table 1. Experiments and results with Jackknife technique. (CFP= Correct Flare Prediction, CFTP= Correct Flare Type Prediction).

Experiment No	Convergence Error	% CFP in Total Flares	% CFTP in Total Flares
1	0.02850	91.50780	88.90815
2	0.03010	93.93414	90.29463
3	0.03400	90.64125	87.69497
4	0.02990	91.68111	90.12132
5	0.02754	89.60139	86.82842
6	0.02755	92.37435	86.82842
7	0.02740	90.29463	86.82842
8	0.04290	92.20104	88.04159
9	0.03070	91.85442	87.86828
10	0.02980	93.41421	89.94801
Average	0.03084	91.75043	88.33622

5 Conclusions

In this paper, an automated NN-based system that provides efficient prediction of solar activities that can affect life on Earth is presented. The system processes two publicly available solar catalogues from the National Geophysical Data Center and compares the reported occurrences of M and X solar flares with the relevant sunspots that were classified earlier and exist in the same NOAA region. To increase the accuracy of prediction, a mathematical model, based on the work of [27], is implemented to simulate the solar activity during the times of flares occurrences. The simulated activ-

ity and classified sunspots are converted to the appropriate numerical formats and fed to cascade forward backpropagation NN, to predict whether a significant flare will occur and whether it is going to be an X or M flare.

Also, with this work, the results found for correct flare prediction in [19] is improved to 91.7% from 85% and correct flare type prediction is improved to 88.3% from 78% by using a simpler topology and also number of inputs in the network is reduced to 4 from 8. By all means this study outperforms the previous work in [19] and can be used for automated flare prediction from sunspot groups.

Our practical findings in this paper show that there is a direct relation between the eruptions of flares and certain McIntosh classes of sunspots such as Ekc, Fki and Fkc. which are in accordance with [33], [34], [35], but it is the first time to verify this relation using machine learning.

We believe that the quality of this work can be enhanced if evolutionary learning is used in conjunction with the learning algorithms presented here. This paper is a first step toward building a real-time flares prediction model. Hence, it is important for our learning algorithms to be able to improve its learning and generalisation capabilities by continuous learning from the new sunspots data, which is available on daily basis. This can be done if automatic learning algorithms that require no user intervention are applied once the prediction model is built. For our future work, we intend on exploring the feasibility of using evolutionary learning, which is slow but very versatile, for our application. The feasibility of designing a hybrid system that combines evolutionary learning with cascade correlation neural networks in a manner similar to Nessy algorithm [36] will be explored. We will also explore the feasibility of integrating SVM in such system. In addition, we intend to apply more testing criteria for the newly developed learning algorithms.

In this work, we have tested our system using the Jack-knife technique, which provides a correct statistical evaluation for the performance of the classifier, when implemented on a limited number of samples. This test has proven that a nonlinear relation exists between sunspots data and the occurrence of flares. However, for our future work we will extend our evaluation criteria by training the model on certain periods of time and testing it on the remaining periods.

References

1. Koskinen, H., E. Tanskanen, R. Pirjola, A. Pulkkinen, C. Dyer, D. Rodgers, and P. Cannon, Space Weather Effects Catalogue, in ESA Space Weather Programme Feasibility Studies, FMI, QinetiQ, RAL Consortium. (2001).
2. Clark, T.D.G., A Review of the Effects of Space Weather on Ground Based Technology, in Space Weather Workshop: Looking towards a European Space Weather Programme: Noordwijk, The Netherlands. (2001).

3. Erinmez, I.A., J.G. Kappenman, and W.A. Radasky, "Management of the geomagnetically induced current risks on the national grid company's electric power transmission system", Journal of Atmospheric and Solar-Terrestrial Physics, 64(5-6): (2002), 743.

4. Barnes, P. and J.V. Dyke, On the Vulnerability of Electric Power to Geomagnetic Storms, Oak Ridge National Laboratory: Tennesse. (1990).

5. Weiher, R. and T. Teisberg, "Economic Valuation of Geomagnetic Storm Forecasts in the North American Electic Industry", Journal of Policy Analysis and Management, 19(2): (2000).

6. Gary, D.E., L.J. Lanzerotti, G.M. Nita, and D.J. Thomson. Effects of Solar Radio Bursts on Wireless Systems. in NATO ARW on Effects of Space Weather on Technology Infrastructure. 25 to 29 March 2003. Rhodes, Greece: Springer, (Year).

7. Shea, M.A. and D.F. Smart, Space weather: The effects on operations in space, in Solar-Terrestrial Relations: Predicting The Effects On The Near- Earth Environment. p. 29-38, (1998).

8. Guy, F. and D.C. Johnson, Space Environmental Impacts on DoD Operations, AIR FORCE SPACE COMMAND. (2003).

9. Worden, S.P., "The Air Force and Future Space Directions: Are We Good Stewards?" Aerospace Power Journal, 15(1): (2001), 50-57.

10. Gao, J.L., H.M. Wang, and M.C. Zhou, "Development of an automatic filament disappearance detection system", Solar Physics, 205(1): (2002), 93-103.

11. Turmon, M., J.M. Pap, and S. Mukhtar, "Statistical pattern recognition for labeling solar active regions: application to SOHO/MDI imagery", Astrophysical Journal, 568(1): (2002), 396-407.

12. Shih, F.Y. and A.J. Kowalski, "Automatic Extraction of Filaments in H-alpha images", Solar Physics, 218(1-2): (2003), 99 – 122.

13. Benkhalil, A., V. Zharkova, S. Ipson, and S. Zharkov. Automatic Identification of Active Regions (Plages) in the Full-Disk Solar Images Using Local Thresholding and Region Growing Techniques. in AISB'03 Symposium on Biologically-inspired Machine Vision, Theory and Application. 2003. University of Wales, Aberystwyth (Year).

14. Lefebvre, S. and J.P. Rozelot, "A new method to detect active features at the solar limb", Solar Physics, 219(1): (2004), 25-37.

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

16. Borda, R.A.F., P.D. Mininni, C.H. Mandrini, D.O. Gomez, O.H. Bauer, and M.G. Rovira, "Automatic solar flare detection using neural network techniques", Solar Physics, 206(2): (2002), 347-357.

17. Qu, M., F.Y. Shih, J. Jing, and H.M. Wang, "Automatic solar flare detection using MLP, RBF, and SVM", Solar Physics, 217(1): (2003), 157-172.

18. Zharkova, V. 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), 148-154.

19. Qahwaji, R. and T. Colak, Neural Network-based Prediction of Solar Activities, in CITSA2006: Orlando. (2006).

20. Liu, C., N. Deng, Y. Liu, D. Falconer, P.R. Goode, C. Denker, and H.M. Wang, "Rapid change of delta spot structure associated with seven major flares", Astrophysical Journal, 622(1): (2005), 722-736.

21. Zirin, H. and M.A. Liggett, "Delta-Spots And Great Flares", Solar Physics, 113(1-2): (1987), 267-283.

22. Shi, Z.X. and J.X. Wang, "Delta-Sunspots And X-Class Flares", Solar Physics, 149(1): (1994), 105-118.

23. National Geophysical Data Center. Available from: ftp://ftp.ngdc.noaa.gov/STP/SOLAR_DATA/ Date Accessed: 2006

24. Holman, G. and S. Benedict, Questions and Answers, NASA. (2006).
25. Hale, G.E., F. Ellerman, S.B. Nicholson, and A.H. Joy, "The Magnetic Polarity of Sun-Spots", Astrophysical Journal, 49: (1919), 153.
26. Greatrix, G.R. and G.H. Curtis, "Magnetic Classification Of Sunspot Groups", Observa-tory, 93(994): (1973), 114-116.
27. Hathaway, D., R.M. Wilson, and E.J. Reichmann, "The Shape of the Sunspot Cycle", Solar Physics, 151: (1994), 177.
28. Horn, D., "Neural Computation Methods and Applications: Summary Talk of the AI", Journal of Nuclear Instruments and Methods in Physics research, 389: (1997), 381-387.
29. Tzafestas, E., A. Nikolaidou, and S. Tzafestas, "Performance Evaluation and Dynamic Node Generation Criteria for ' Principal Component Analysis' Neural Network", Mathemat-ics and Computer in Simulation, 51: (2000), 145-156.
30. Ma, Q., A.Yan, Z. Hu, Z. Li, and B. Fan, "Principal Component Analysis and Arti-ficial Neural Networks Applied to the Classification of Chinese Pottery Neolithic age", Analytica Chimica Acta Journal, 406: (2000), 247-256.
31. Kim, J., A. Mowat, P. Poole, and N. Kasabov, "Linear And Non-Linear Pattern Recogni-tion Models For Classification Of Fruit From Visible-Near Infrared Spectra", Chemometrics And Intelligent Laboratory Systems, 51: (2000),201-216.
32. Fukunaga, K., Introduction to Statistical Pattern Recognition," Academic Press, New York, 1990. New York: Academic Press
(1990).
33. McIntosh, P.S., "The Classification Of Sunspot Groups", Solar Physics, 125(2): (1990), 251-267.
34. Warwick, C.S., " Sunspot Configurations and Proton Flares ", Astrophysical Journal, 145: (1966), 215.
35. Sakurai, K., "On the magnetic configuration of sunspot groups which produce solar proton flares", Planetary and Space Science, 18(1): (1970), 33.
36. Köppen, M., M. Teunis, and B. Nickolay. Nessy - An Evolutionary Learning Neural Net-work. in SOCO 97. 1997. Nimes, France(Year).