Journal of Theoretical and Applied Information Technology
10th April 2015. Vol.74 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

SOLAR FLARE M-CLASS PREDICTION USING ARTIFICIAL INTELLIGENCE TECHNIQUES

[1*]AZAM ZAVVARI, [2]MOHAMMAD TARIQUL ISLAM, [3]RADIAL ANWAR, [4]ZAMRI ZAINAL ABIDIN

[1]Department of Electrical, Electronic and Systems Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

[2] Prof., Department of Electrical, Electronic and Systems Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

[3]Space Science Center, Institute of Climate Change, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

[4]Dr., Department of Physics, Faculty of Science Building, University of Malaya, 50603 Kuala Lumpur, Selangor, Malaysia

E-mail: [1]azam.zavvari@gmail.com*, [2]titareq@gmail.com, [3]adi_astro@yahoo.com, [4]zzaa@um.edu.my

ABSTRACT

Currently, astronomical data have increased in terms of volume and complexity. To bring out the information in order to analyze and predict, the artificial intelligence techniques are required. This paper aims to apply artificial intelligence techniques to predict M-class solar flare. Artificial neural network, support vector machine and naïve bayes techniques are compared to define the best prediction performance accuracy technique. The dataset have been collected from daily data for 16 years, from 1998 to 2013. The attributes consist of solar flares data and sunspot number. The sunspots are a cooler spot on the surface of the sun, which have relation with solar flares. The Java-based machine learning WEKA is used for analysis and predicts solar flares. The best forecasted performance accuracy is achieved based on the artificial neural network method.

Keywords: *Neural Network, Support vector machine, Naïve Bayes, Solar Flare, Artificial Intelligence Techniques.*

1. INTRODUCTION

Solar activities have a great impact on mankind technological systems as well as life on the Earth. For instance, solar flares and coronal mass ejections, two types of solar activities which may have a huge impact on communication and navigation satellites by degrading satellite signals and reducing their accuracy. Moreover, the satellites can be at risk of space weather effects [1].

A flare is represented as an intense, rapid eruptions and sudden variation in brightness. The five different types of solar flares are defined as follows. The lower classes of solar flare are the A-class and B-class among the rest of other classes. These classes happen commonly during solar minimum and maximum, which are not involved in this study. C-Class is known as small solar flares and sometimes produces the slow Coronal Mass Ejection (CME). This class has almost no effect on earth. Medium large solar flares which called M-class may cause radio blackout and radiation storms. The X-class is a strongest solar flare can cause radio blackout on the daylight side of the Earth, strong geomagnetic storming and also Coronal Mass Ejection [1]. Another element which has close relation with solar flares is sunspots. Sunspots are the dark spots on the sun's surface; they are dark because these spots are cooler than the surrounding area. The temperature of this cool area is approximately 1500 K. The biggest sunspot which ever observed was covering approximately 50,000 km of the sun's surface. This large sunspot was big enough to be observed with the naked eye.

Due to the importance of solar activity, numerous researches have been conducted to monitor the Sun. Nowadays a large number of telescopes and antennas are used to observe solar activities. Researchers are building telescopes to gain data that producing terabytes of data each year. In the past decades, astronomical data volumes have grown from gigabytes into terabytes, and these volumes

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

will continue to grow from terabytes to petabytes in the recent decade. This exponential growth of astronomical data needs new artificial intelligence techniques for large databases. The traditional process of knowledge discovery does not have good performance in large databases; therefore a new computational approach is needed. Due to the large data volume and complexity of data, artificial intelligence becomes the only hope for extracting knowledge [2].

The Sun is the most powerful energy driving source and solar flare is one of the most important and intense solar activity phenomenons. Solar activities can effects life on earth and may disturb communication and power systems and hence it is important to predict solar flare accurately. Solar flare forecasting is facing many challenges such as automated definition, classification and prediction.

The paper is structured as follows: section 2 reviews a related study of different methods which applied for prediction solar flares. Then, three artificial intelligence methods are chosen and defined. In the next section, details about the solar flare dataset and attributes are explained. In section 5, a Java based WEKA machine learning tool is applied for analysis and prediction of solar flare data.

2. BACKGROUND OF STUDY

Prediction methods are used to protect the astronauts and spacecraft from radiation and high-energy particles of a big solar flare. The nature of solar feature is not well-defined yet, which make automated forecast become difficult. Many statistical and artificial intelligence methods employ the observational data to establish a high accuracy solar flare prediction. The traditional process of knowledge discovery such as statistical methods does not have good performance in large databases; therefore a new computational approach is needed. In recent years, more and more artificial intelligence methods have been applied for flare prediction in different solar activities dataset and data types with dissimilar time duration.

Bradshaw et al. (1989) built a three-layer back-propagation neural network where the system took solar data as an input and predict the flare occurrence. The solar data was collected from THEO database [3]. Borda et al. (2002) described a new automatic solar events detection used the neural network multi-layer perceptron and back-propagation. This method applied on image data which is obtain from HASTA full-disk solar images [4]. Qu et al. (2003) compared three classification methods for solar flares features extracted. Support vector machines, multi-layer perceptron neural network and radial basis function was three experimented methods which utilizes the solar Hα data from Big Bear Solar Observatory [5].

Qahwaji and Colak (2007) applied cascade correlation neural network, support vector machines and radial basis function networks for a short-term solar flare prediction. The system attempts to find a relation between solar cycle data and McIntosh sunspot classification [6]. Li et al. (2007) proposed a solar flare forecasting method with the combined method of support vector machines and k-nearest neighbor algorithm. The method uses the 10 cm radio flux and sunspots for predicting M class flare [7]. Wang et al. (2008) proposed a solar flare forecasting model based on artificial neural network on solar magnetic field data [8]. Qahwaji et al. (2008) used a machine learning system to predict the flare with their associated CMEs. Two SOHO/LASCO CME and NGDC flares catalogues was used as input for the system. [9]. Mubiru (2011) developed an Artificial neural network for predicting monthly average solar irradiation at four same climate locations [10]. Bian et al. (2013) Extreme Learning Machine as a fast learning algorithm present enhanced generalization performances for flare forecasting methods. The parameter related to this study was three magnetic parameters and magnetic energy dissipation [11].

3. ARTIFICIAL INTELLIGENCE TECHNIQUES

This section reviews ANN, SVM and Naïve Bayes techniques which are choose based on literature. The methods are defined as follows:

The ANN is a type of artificial intelligence, which have the ability to model linear and nonlinear processing. As opposed to the traditional modeling techniques, ANN is a data driven, black-box and self-adaptive method. ANN is applied to various fields of science and solar activities such as estimated global solar radiation and empirical regression. ANN is one of the most commonly used techniques which have wider application in the solar activity area. This method has been proved as a powerful tool in data analysis and data processing system where a large number of data is available.

ANN is an artificial representation of the human brain that aims to simulate its learning process. The brain consists of four main units. Dendrites are interconnected as a set of nerve cells, or

Journal of Theoretical and Applied Information Technology
10th April 2015. Vol.74 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

information-processing unit. The nucleus, which is a core of the cell body, and a single long fiber called the axon. The axon can communicate with another neuron with synapses.

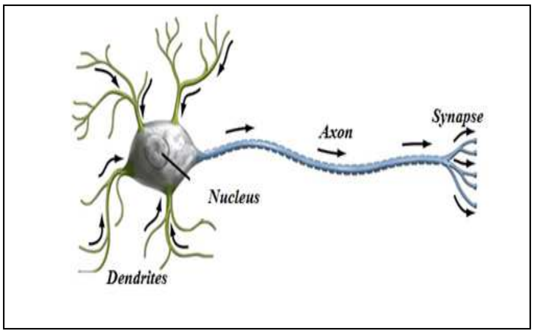

Figure 1. Biological View Of Neuron.

ANN as one of the most powerful approximation multidimensional functions, have very wide applications in machine learning such as prediction, pattern recognition, and signal de-noising and so on. Based on simplistic biological neuron model ANN is defined as follows:

Artificial neuron receives a number of inputs as a (x_1, x_2,…, x_n) and each of them multiple it own weight (w_1, w_2,…, w_n) and it sums with an additional bias number (see Equation (1)):

$$x_1 w_1 + x_2 w_2 + \ldots + x_n w_n + \mathrm{b} = \sum w_\mathrm{i}\, x_\mathrm{i} + \mathrm{b} \quad (1)$$

After that it goes through some nonlinear function. Figure 2 illustrates the information process in a neuron.

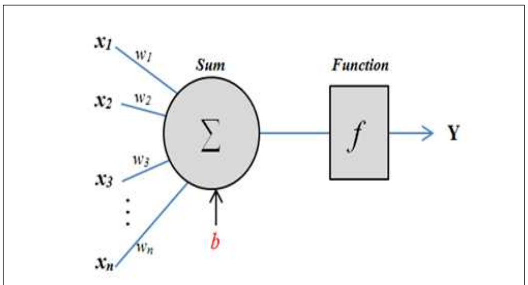

Figure 2. Information Processing With A Neuron.

The neural network system has three different layers. The first layer is called an input layer and the last layer is the output layer, and the rest of layers between these two layers called hidden layers. A network of neurons with three hidden layers is shown in Figure 3. The entrance of the first hidden layer received the value of the input layer (x_i) multiplied with the weight of that specific input. After that a bias value is added for each neuron then a function is applied, which is an Equation (2), and it happens again for the second hidden layer.

$$\varphi\ (w^\mathrm{T} x_i + b) \quad (2)$$

The second hidden layer entrance is the resulting output of the first hidden layer (see Equation (3)):

$$\varphi\ (v^\mathrm{T}\ \varphi\ (w^\mathrm{T} x_i + b) + c) \quad (3)$$

And in the end, the third hidden layer value will add to the network and the Equation (4) is generated:

$$\mathrm{Y} = d^\mathrm{T}\ \varphi\ (v^\mathrm{T}\ \varphi\ (w^\mathrm{T} x + b) + c) \quad (4)$$

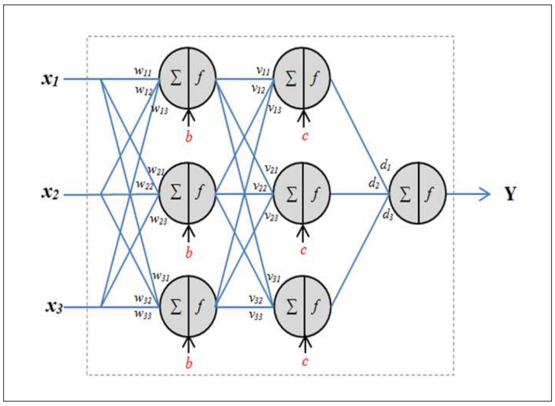

Figure 3. Neural Network System With Three Hidden Layers.

The Support Vector Machines (SVM) is a powerful data analysis algorithm applies for classification and regression analysis. The SVM models have similar functionality to Artificial Neural Networks and Radial Basis Function. It can be applied for modeling the complex and real-world problems.

The Naïve Bayes is a classifier algorithm based on Bayesian theorem. This algorithm is most suitable when the input dimensionality is high. Despite the fact that the naïve bayes algorithm is simple, but can often outperform more complicated classification methods.

4. DATASET PREPARATION

The dataset is collected from space weather prediction center *(http://www.swpc.noaa.gov)*. It is a dataset collected for 16 years from 1998 till 2013.

Journal of Theoretical and Applied Information Technology
10th April 2015. Vol.74 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

In this part, the M-class solar flare is used as a class label and change into negative and positive type of data. The M-class data which is a class data have negative type (Mean: No solar flare M-class happen in that specific day) and positive data (Mean: at least one M-class happen in the specific day). All the data are provided in CSV format four inputs and one class data. The attributes of datasets are different types C-class and X-class of solar flares and sunspot number. M-class data have 4827 days negative and 1017 days positive data for 16 years. Table 1 shows a statistical analysis of all the attributes in the dataset. The results are performed by using Java based WEKA machine learning.

Table 1. Statistical Attribute Analysis.

Row ID	Sunspot Number	Xray C-class	Xray X-class
Minimum	0	0	0
Maximum	401	26	3
Mean	81.569	2.887	0.026
Std. Deviation	69.715	3.798	0.184
Variance	4860.11	14.423	0.034
Overall sum	476689	16873	150
No. Missing	0	0	0
Row count	5844	5844	5844

1. EVALUATION MEASUREMENTS

Classification technique assigns different type of data for collection to different classes or categories. The aim of classification technique is to predict the needed data. Four important categories for M-class solar flare which are used in the rest of classification method are explained here.

- The negative samples that are classified correctly as negative are called True Negative (TN).
- The positive samples that are classified correctly as positive are called True Positive (TP).
- The positive samples that are classified incorrectly as negative are called False Negative (FN).
- The negative samples that are classified incorrectly as positive are called False Positive (FP).

These rates are given by Equations. The confusion matrix reports the validity of the classification models. The quality and validity measures of the classification are built on the confusion matrix by WEKA. Table 2-4 shows confusion matrixes of artificial neural network, support vector machine and naïve bayes classifiers.

Table 2. Confusion Matrix Of Artificial Neural Network Method.

Method	ANN	
Class	True	False
Positive	TP = 0.268	FP = 0.05
Negative	TN = 0.95	FN = 0.732

Table 3. Confusion Matrix of Support vector machine method.

Method	SVM	
Class	True	False
Positive	TP = 0.071	FP = 0.009
Negative	TN = 0.991	FN = 0.929

Table 4. Confusion Matrix Of Naïve Bayes Method.

Method	NB	
Class	True	False
Positive	TP = 0.48	FP = 0.108
Negative	TN = 0.892	FN = 0.52

Classifier Accuracy Evaluations is calculated based on confusion Matrix. F-value are measured based on precision and recall value. The accuracy, F-value, precision and recall of the classifier is determined by Equations (5-8).

$$Classifier\ \ Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FN} + \mathrm{FP} + \mathrm{TN}} \tag{5}$$

$$\mathrm{Precision} = \frac{TP}{TP + FP} \tag{6}$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{7}$$

$$F - value = 2 \times \frac{precision \times recall}{precision + recall} \tag{8}$$

The data are analyzed by WEKA software based on 10-fold cross-validation. Cross-validation is a technique for performance estimation. This technique partitioned the data into 10 folds. All folds have the same size. The method is applied ten

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

times with different test and training set and the average of accuracy are chosen as a final accuracy result.

Table 5 shows that the 10-fold cross-validation accuracy of the ANN is 83.1451 percent, which is a high accuracy. SVM accuracy is lower than ANN but this method has lowest classification errors.

Table 5. Comparison Of The Discussed Methods.

Method	ANN	SVM	NB
Precision	0.842767	0.8875	0.816326531
Recall	0.268	0.071	0.48
F-Value	0.406677	0.131481481	0.604534005
Accuracy	83.1451%	83.0595%	82.0671%
Error	0.2307	0.1694	0.2028

5. CONCLUSION

Solar feature forecasting is an emerging science that facing many challenges, because the nature of solar features is not clear yet. To find the best method with the good accuracy, three artificial intelligence methods are applied on solar flare dataset. The multi-layer perceptron neural network provides efficient accuracy compared to other methods while support vector machine has been proven that it has the lowest error among the investigated methods.

ACKNOWLEDGMENTS:

The authors would like to thank the Space Science Center (ANGKASA), Universiti Kebangsaan Malaysia; and National Space Agency of Malaysia research grant LL-2012-010; and University of Malaya's HIR grant UM.S/625/3/HIR/28.

REFERENCE:

[1] T. Colak, R. Qahwaji, "Automated Solar Activity Prediction: A Hybrid Computer Platform Using Machine Learning and Solar Imaging for Automated Prediction of Solar Flares", Space weather, Vol. 7, No. 6, Jun 2009, S06001, DOI: 10.1029/2008SW000401.

[2] K.D. Borne, "Astroinformatics: A 21st Century Approach to Astronomy", Instrumentation and Methods for Astrophysics Sep 2009, pp.1-14. http://arxiv.org/pdf/0909.3892v1.pdf.

[3] G. Bradshaw, R. Fozzard, L. Ceci, "A connectionist expert system that actually works", Advances in neural information processing systems, 1989, pp. 248-255.

[4] R.A.F. Borda, P.D. Mininni, C.H. Mandrini, D.O. Gómez, O.H. Bauer, M.G. Rovira, "Automatic solar flare detection using neural network techniques," Solar Physics, Vol. 206, 2002, pp. 347 – 357. Doi:10.1023/A:1015043621346.

[5] M. Qu, F.Y. Shih, J. Jing, H. Wang, "Automatic solar flare detection using MLP, RBF, and SVM" Solar Physics, Vol. 217, 2003, pp. 157 – 172. Doi:10.1023/A:1027388729489.

[6] R. Qahwaji, T. Colak, "Automatic short-term solar flare prediction using machine learning and sunspot associations", Solar Physics, Vol. 241, No. 1, 2007, pp. 195 – 211.

[7] R. Li, H. N. Wang, H. He, Y. M. Cui, Z.L. Du, "Support Vector Machine combined with K-Nearest Neighbors for Solar Flare Forecasting", Chinese Journal of Astronomy and Astrophysics, Vol. 7, No. 3, 2007, pp. 441-447.

[8] H. N. Wang, Y. M. Cui, R. Li, L. Y. Zhang, H. Han, " Solar flare forecasting model supported with artificial neural network techniques", Advances in Space Research, Vol. 42, 2008, pp. 1464-1468.

[9] R. Qahwaji, T. Colak, M. Al-Omari, S. Ipson, "Automated Machine Learning-Based Prediction of CMEs Based on Flare Associations", Solar Physics, Vol. 248, No. 2, 2008, pp. 471-483.

[10] J. Mubiru, "Using Artificial Neural Networks to Predict Direct Solar Irradiation", Advances in Artificial Neural Systems, Vol. 2011, 2011.

[11] Y. Bian, J. Yang, M. Li, R. Lan, "Automated Flare Prediction Using Extreme Learning Machine", Mathematical Problems in Engineering, 2013.