Drought Monitoring: A Performance Investigation of Three Machine Learning Techniques

Pheeha Machaka(✉)

School of Computing, University of South Africa, Science Campus, Florida Park, Johannesburg 1709, South Africa
machap@unisa.ac.za

Abstract. This paper investigates the use of Soft Computing techniques on a drought monitoring case study. This is in effort to create an intelligent middleware for Ubiquitous Sensor Networks (USN) using machine learning techniques. Algorithms in Artificial Immune System, Neural Networks and Bayesian Networks were used. The paper reveals the results from an experiment on data collected over 95 years in the Trompsburg region of the Free State Province, South Africa.

Keywords: Machine learning · Algorithms · Neural networks · Artificial immune systems · NaiveBayes · Standard precipitation index

1 Introduction

Drought is a natural, environmental disaster that can be classified together with earthquakes, epidemics and floods. It has substantial impact on humans and its impact can persist for several years. A current example of an extreme drought case is that of East Africa. East Africa has experienced its worst drought in 60 years, affecting more than 11 million people. It was declared as a famine-stricken region, with overcrowded refugee camps in Kenya and Ethiopia. Livestock is dying at a rapid rate, with one farmer reporting have lost 17 goats in one day. Officials also warn that over 800 000 children could die from malnutrition across East Africa nations of Somalia, Ethiopia, Eritrea and Kenya.

Given the impacts of drought, there is a need for developing strategies for drought monitoring and early warning systems that are able to determine drought severity. These systems can help in planning and managing water resources, and can help reduce and avoid impacts of drought.

As part of our drought monitoring research study in our research group, Masinde and Bagula in [14] proposed a drought prediction framework that combines mobile phones and wireless sensor networks to be able to capture and relay micro drought parameters. The framework is an enhancement of ITU's Ubiquitous Sensor Network (USN) Layers. Ubiquitous Sensor Networks are described in [9] as networks of intelligent sensor nodes that could be deployed anywhere, anytime, by anyone and anything. The framework is further described as composing 5 layers: **Sensor**

P.C. Vinh et al. (Eds.): ICCASA 2013, LNICST 128, pp. 47–56, 2014.
DOI: 10.1007/978-3-319-05939-6_5,

Networking, USN Access Network, Network Infrastructure, USN Middleware, and USN Applications Platform.

Our focus in this research study is on the 4th layer, the USN Middleware. This is composed of intelligent software that will help with drought monitoring and prediction. In this study we investigate available drought monitoring tools, and look into the use of learning algorithms in drought forecasting. Three learning algorithms, viz. Artificial Immune Systems, Bayesian Networks and Artificial Neural Networks, are studied and their performance on a South African precipitation dataset is compared.

2 Background and Related Work

2.1 Drought Indices

Research in previous years has developed indices that measure drought. These indices can be used in early warning and drought monitoring systems. These indices are very important in measuring the drought severity, intensity, duration, coverage and magnitude. Mishra and Singh in [18] made great attempts to make comparisons and find out which of the drought indices are most suitable for drought monitoring. In their research they listed the relevant indices and evaluated them according to regions were the indices are used; the type of drought being monitored; how indices are used; advantages and disadvantages; and the overall general usefulness.

The comparison between SPI (Standard Precipitation Index) and PDSI (Palmer Drought Severity Index) came out with the following results:

1. "SPI is more representative of short-term precipitation than PDSI and thus is a better indicator for soil moisture variation and soil wetness" [20].
2. "SPI provides a better spatial standardization than does PDSI with respect to extreme drought events" [12].
3. It was found that the SPI was a valuable estimator for drought severity [18].
4. SPI detects the onset of a drought earlier than PDSI [7].

Based on this, it can be inferred that the SPI is a better monitoring tool to use. We will therefore focus on the SPI for the remainder of this case study.

2.2 Algorithms

There are various algorithms that exist with different variations for the three chosen methods. The algorithms that were used for this paper's experiments, were those found in the WEKA libraries [21].

- **Artificial Neural Networks.** ANN's are mathematical or computational models that get their inspiration from biological neural systems. In this paper the neural network model, Multilayer Perceptron (MLP) was used to conduct experiments. The MLP is a feed forward neural network model in which vertices are arranged in layers. MLP have one or more layer(s) of hidden nodes, which are not directly connected to the input and output nodes [5]. For the purpose of this experiment we employed Weka's Multilayer Perceptron implementation.

- **Bayesian Networks.** Bayesian Networks can be described briefly as Acyclic Directed Graph (DAG) which defines a factorisation of a joint probability distribution over the variables that are represented by the nodes of the DAG, where the factorisation is given by the direct links of the DAG [11]. The NaiveBayes algorithm was used for the experiments. It makes a strong assumption that all attributes of the examples are independent of each other given the context of the class. The Weka's NaiveBayes implements this probabilistic Naïve Bayes classifier [23].
- **Artificial Immune Systems.** The AIS takes inspiration from the robust and powerful capabilities of the Human Immune System's (HIS) capabilities to distinguish between self and non-self [13]. The Algorithm employed in this paper's experiments is the Weka's Artificial Immune Recognition System (AIRS) learning algorithm [22]. The AIRS is a supervised AIS learning algorithm that has shown significant success on a broad range of classification problems [3, 6, 13].

2.3 Related Work

There has been creditable work done to predict weather condition using Bayesian Networks, and in [10] they were applied to the problem of predicting sea breeze. Bayesian Networks were then compared with existing rule-based system and it was found out that the Bayesian network outperformed the traditional rule-based system in prediction accuracy.

Authors in [2] introduce a Bayesian Network framework that deals with multivariate spatiality distributed time series. They used it to predict precipitation for 100 stations in the North basin of the Iberian Peninsula during winter of 1999. In [4], Bayesian networks are used to estimate forecasts of peak and average temperatures. In this case study, data derived from a power utility system is used to forecast electric load with imperfect information.

Considerable weather forecast work has focused on the use of Artificial Neural Networks (ANN) and Bayesian Networks, but only a few use artificial immune systems for weather forecasting. Authors in [23] implemented an immune-based algorithm that was applied on weather data for forecasting. The immune algorithm was compared to an artificial neural network algorithm and the results reveal that the implemented immune algorithm had a higher forecast accuracy rate than that of the neural network.

There has also been great work done in the field of using artificial neural networks in drought monitoring. Antonic et al. [1] used feed-forward ANN with Multilayer Perceptron (MLP) for empirical model development using seven climatic variables (monthly mean air temperature, monthly mean daily minimum and maximum air temperature, monthly mean relative humidity, monthly precipitation, monthly mean global solar irradiation and monthly potential evapotranspiration).

Authors in [17] used a record of SPI time series data and linear stochastic models, recursive multistep neural networks (RMSNN) for drought forecasting in the Kangsabati river basin, which lies in the Purulia district of West Bengal, India. In their comparison they found neural networks to be more suitable for drought forecasting. Sajikumar [19] used a Temporal Back-Propagation Neural Network (TBP-NN) for monthly rainfall-runoff modeling in scarce data conditions.

In this study, we would like to investigate the performance of the three learning algorithms in the aim to answer the following questions:

- Which method performs better?
- How do the methods perform across different SPI time scales?
- What kind of mined data is extracted using the methods?

3 Research Design

The section that follows will describe the methods and techniques used to carry out the research presented in this paper.

3.1 Drought Monitoring Region and Data Collection

The region of monitoring is Trompsburg, Free State, South Africa. Trompsburg is a small town located in the southern Free State. It is in the ecotone between Nama-Karoo and the grassland biome. The main land use in the region is livestock farming, especially sheep and cattle farming. Monthly precipitation data was collected from this region by the South African Weather Services for the period January 1913 to May 2009; making a total of 96 years of observations.

3.2 Algorithms Performance Evaluation

To measure performance of the algorithms, the following accuracy measures were used:

- **True Positive Rate.** This refers to the function of true positives out of the positives.
- **False Positive Rate.** This refers to the function of false positives out of the positives.
- **Kappa Statistic.** This is used to measure the success of a predictor, the agreement between predicted and observed categorisation of a dataset, while correcting for agreement that occurs by chance [23].
- **F-Measure.** There is a trade-off between Precision and Recall measures. When one tries to improve the first measure, there is often deterioration in the second measure. The F-measure provides a harmonic mean precision and recall.

3.3 Test Cases

In the literature reviewed above, it was found out that there are four different types of drought: meteorological, hydrological, agricultural, and socio-economic droughts.

Using Standard Precipitation Index (SPI) allows for monitoring the different types of drought, by using different time scales [8, 15–18].

This case study will focus on the following time scales: SPI 3 months, SPI 6 months, SPI 12 months and SPI 24 months for the Trompsburg region.

3.4 Experiment Design

Precipitation data was used for the Trompsburg region. We had to transform and calculate the data into precipitation data in such a way that we can create training and testing dataset. The steps taken for completing the experiments in this study are shown below by Fig. 1.

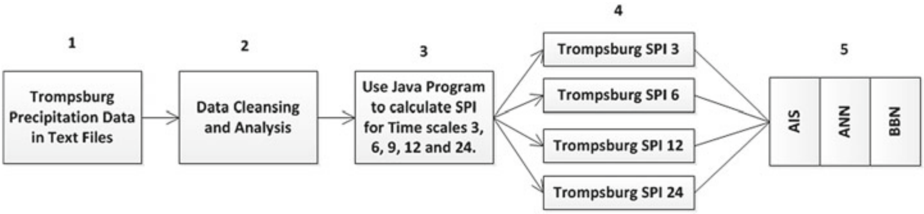

Fig. 1. Trompsburg experiment

1. The data is prepared into text files for further processing.
2. Data cleansing and Analysis: Data is checked for any missing data. If there is missing data, the average precipitation is used.
3. A Java program is designed for calculating the monthly SPI when given time-scale inputs and monthly precipitation data. This program can be found and downloaded from (http://greenleaf.unl.edu/downloads/).
4. The files are output produced by the Java Program for time scales 3, 6, 12 and 24.
5. The data is then used by the Weka program to test performance measures using the AIS, ANN and BNN algorithms. A 10-fold cross validation technique was used, and for each of the four test cases, the experiments were iterated 10 times. The mean algorithm performance measures of each experiment's iteration were recorded and used for statistical comparisons.

The results for the Trompsburg drought experiment follows in the section below. In the evaluation the t-test statistics are used.

4 Results and Discussion

The results of the research are discussed below and algorithms are assessed based on the following algorithm performance measures: Kappa Statistic, True Positive Rate, False Positive Rate and F-measure.

4.1 True Positive Rate Performance

The MLP had an average true positive rate of 68.58 %, while the NaiveBayes' was slightly lower at 68.75 %. The AIRS2 had an even lower average true positive rate performance of 55.47 %.

For the test:

$$\begin{aligned}&\text{True Positive Rate: } H_0\text{: } \mu_{\mathrm{MLP}} - \mu_{\mathrm{NaiveBayes}} = 0\\&\text{True Positive Rate: } H_1\text{: } \mu_{\mathrm{MLP}} - \mu_{\mathrm{NaiveBayes}} \neq 0\end{aligned} \tag{1}$$

In Table 1, the value of the t-Statistic is −2.5572 and its two-tailed p-value is 0.0109. At the 5 % confidence level, the test is significant and there is strong evidence to infer that the alternative hypothesis is true. Therefore we reject the null hypothesis and conclude that there is a difference in the mean true positive rate for the MLP and NaiveBayes algorithms.

For the test:

$$\begin{aligned}&\text{True Positive Rate: } H_0\text{: } \mu_{\mathrm{MLP}} - \mu_{\mathrm{AIRS2}} = 0\\&\text{True Positive Rate: } H_1\text{: } \mu_{\mathrm{MLP}} - \mu_{\mathrm{AIRS2}} \neq 0\end{aligned} \tag{2}$$

In Table 1, the value of the t-Statistic is 60.5019 and its two-tailed p-value is 0.0000. At the 5 % confidence level, the test is significant and there is overwhelming evidence to infer that the alternative hypothesis is true. Therefore we reject the null hypothesis and conclude that there is a difference in mean true positive rate for the MLP and AIRS2 algorithms.

4.2 False Positive Performance

The average false positive rate for the MLP was 68.21 % across all SPI test cases, while that of the NaiveBayes was slightly lower at 68.61 %. The AIRS2 had an impressive lower average false positive rate performance of 54.82 %.

For the test:

$$\begin{aligned}&\text{False Positive Rate: } H_0\text{: } \mu_{\mathrm{MLP}} - \mu_{\mathrm{NaiveBayes}} = 0\\&\text{False Positive Rate: } H_1\text{: } \mu_{\mathrm{MLP}} - \mu_{\mathrm{NaiveBayes}} \neq 0\end{aligned} \tag{3}$$

Table 1. Results for true positive rate t-test for paired two samples for means

t-Test: Paired Two Sample for Means					
	Multilayer Perceptron	*NaiveBayes*		*Multilayer Perceptron*	*AIRS2*
Mean	0.6859	0.6878	Mean	0.6859	0.5548
Variance	0.0003	0.0002	Variance	0.0003	0.0015
Observations	400	400	Observations	400	400
Hypothesized Mean Difference	0		Hypothesized Mean Difference	0	
t Stat	-2.5572		t Stat	60.5019	
P(T<=t) two-tail	0.0109		P(T<=t) two-tail	0.0000	

Table 2. Results for false positive rate t-test for paired two samples for means

t-Test: Paired Two Sample for Means					
	Multilayer Perceptron	*NaiveBayes*		*Multilayer Perceptron*	*AIRS2*
Mean	0.6821	0.6862	Mean	0.6821	0.5482
Variance	0.0006	0.0004	Variance	0.0006	0.0026
Observations	400	400	Observations	400	400
Hypothesized Mean Difference	0		Hypothesized Mean Difference	0	
t Stat	-4.0369		t Stat	50.5275	
P(T<=t) two-tail	0.0001		P(T<=t) two-tail	0.0000	

In Table 2, the value of the t-Statistic is −4.0369 and its two-tailed p-value is 0.0001. At the 5 % confidence level, the test is significant and there is overwhelming evidence to infer that the alternative hypothesis is true. Therefore we reject the null hypothesis and conclude that there is a difference in the mean false positive rate for the MLP and NaiveBayes algorithms.

For the test:

$$\begin{aligned} &\text{False Positive Rate: } H_0\text{: } \mu_{\text{MLP}} - \mu_{\text{AIRS2}} = 0 \\ &\text{False Positive Rate: } H_1\text{: } \mu_{\text{MLP}} - \mu_{\text{AIRS2}} \neq 0 \end{aligned} \tag{4}$$

In Table 2, the value of the t-Statistic is 50.5275 and its two-tailed p-value is 0.0000. At the 5 % confidence level, the test is significant and there is overwhelming evidence to infer that the alternative hypothesis is true. Therefore we reject the null hypothesis and conclude that there is a difference in mean false positive rate for the MLP and AIRS2 algorithms.

4.3 Kappa Statistic Performance

Across all SPI test cases, the MLP had an average Kappa statistic of 0.51 %, while that of the NaiveBayes was 0.27 %. The AIRS2 had an average kappa statistic of 0.67 %.

For the test:

$$\begin{aligned} &\text{Kappa Statistic: } H_0\text{: } \mu_{\text{MLP}} - \mu_{\text{NaiveBayes}} = 0 \\ &\text{Kappa Statistic: } H_1\text{: } \mu_{\text{MLP}} - \mu_{\text{NaiveBayes}} \neq 0 \end{aligned} \tag{5}$$

In Table 3, the value of the t-Statistic is 1.865 and its two-tailed p-value is 0.0628. At the 5 % confidence level, the test is not statistically significant and there is weak evidence to infer that the alternative hypothesis is true. Therefore we do not reject the null hypothesis and conclude that the difference in the mean Kappa statistic for the MLP and NaiveBayes algorithms equals zero, the hypothesised mean.

Table 3. Results for Kappa statistic t-test for paired two samples for means

t-Test: Paired Two Sample for Means					
	Multilayer Perceptron	*NaiveBayes*		*Multilayer Perceptron*	*AIRS2*
Mean	0.0052	0.0028	Mean	0.0052	0.0068
Variance	0.0006	0.0004	Variance	0.0006	0.0033
Observations	400	400	Observations	400	400
Hypothesized Mean Difference	0		Hypothesized Mean Difference	0	
t Stat	1.8658		t Stat	-0.5169	
P(T<=t) two-tail	0.0628		P(T<=t) two-tail	0.6055	

For the test:

$$\begin{aligned} &\text{Kappa Statistic: } H_0\text{: } \mu_{MLP} - \mu_{AIRS2} = 0 \\ &\text{Kappa Statistic: } H_1\text{: } \mu_{MLP} - \mu_{AIRS2} \neq 0 \end{aligned} \tag{6}$$

In Table 3, the value of the t-Statistic is -0.5169 and its two-tailed p-value is 0.6055. At the 5 % confidence level, the test is not statistically significant and there is little to no evidence to infer that the alternative hypothesis is true. Therefore we do not reject the null hypothesis and conclude that the difference in mean Kappa statistic for the MLP and AIRS2 algorithms equals zero, the hypothesised mean.

4.4 F-Measure Performance

The MLP had an average SPI F-measure performance of 56.86 %, and that of the NaiveBayes was slightly lower at 56.79 %. The AIRS2 algorithm had a lower average SPI F-measure performance of 52.53 %.

For the test:

$$\begin{aligned} &\text{F-Measure: } H_0\text{: } \mu_{MLP} - \mu_{NaiveBayes} = 0 \\ &\text{F-Measure: } H_1\text{: } \mu_{MLP} - \mu_{NaiveBayes} \neq 0 \end{aligned} \tag{7}$$

In Table 4, the value of the t-Statistic is 1.4567 and its two-tailed p-value is 0.1467. At the 5 % confidence level, the test is not statistically significant and there is

Table 4. Results for F-measure t-test for paired two samples for means

t-Test: Paired Two Sample for Means					
	Multilayer Perceptron	*NaiveBayes*		*Multilayer Perceptron*	*AIRS2*
Mean	0.5686	0.5679	Mean	0.5686	0.5254
Variance	0.0003	0.0002	Variance	0.0003	0.0009
Observations	400	400	Observations	400	400
Hypothesized Mean Difference	0		Hypothesized Mean Difference	0	
t Stat	1.4567		t Stat	28.6235	
P(T<=t) two-tail	0.1460		P(T<=t) two-tail	0.0000	

little or no evidence to infer that the alternative hypothesis is true. Therefore we do not reject the null hypothesis and conclude that the difference in the mean recall for the MLP and NaiveBayes algorithms equals zero, the hypothesised mean.

For the test:

$$\begin{aligned} &\text{F-Measure: } H_0\text{: } \mu_{MLP} - \mu_{AIRS2} = 0 \\ &\text{F-Measure: } H_1\text{: } \mu_{MLP} - \mu_{AIRS2} \neq 0 \end{aligned} \tag{8}$$

In Table 4, the value of the t-Statistic is 28.6235 and its two-tailed p-value is 0.0000. At the 5 % confidence level, the test is significant and there is overwhelming evidence to infer that the alternative hypothesis is true. Therefore we reject the null hypothesis and conclude that there is a difference in mean recall for the MLP and AIRS2 algorithms.

5 Conclusions

The statistical experiments conducted above for algorithm performance measures indicate that the mean Kappa statistic for the MLP, NaiveBayes and AIRS2 algorithms were statistically similar. The mean F-measure for the MLP and NaiveBayes were also statistically similar.

Overall, across all performance measures, one can then safely conclude that there was a significant difference in mean algorithm performance measures for the MLP, NaiveBayes and AIRS2 algorithms. The MLP had a better performance than the NaiveBayes and the AIRS2 algorithms. If applied to the drought case study, the MLP will produce better results. The average rate for most performance measures was below 65 % and in some cases, 15 %. But when one carefully examines the results, one finds out that the three algorithms had a mediocre to poor classification performance.

References

1. Antonic, O., Krizan, J., Marki, A., Bukovec, D.: Spatio-temporal interpolation of climatic variables over large region of complex terrain using neural networks. Ecol. Modelling **138**, 255–263 (2001)
2. Antonio, S.C., Rafael, C., Carmen, S., Jose, M.G.: Bayesian networks for probabilistic weather prediction. In: Anonymous, pp. 695–700 (2002)
3. Dasgupta, D., Gonzales, F.: An immunity-based technique to characterize intrusions in computer networks. IEEE Trans. Evol. Comput. **6**, 281–291 (2002)
4. Douglas, A.P., Breilphol, A.M., Lee, F.N., Adapa, R.: The impacts of temperature forecast uncertainty on Bayesian load forecasting. IEEE Trans. Power Syst. **13**, 1507–1513 (1998)
5. Dunne, R.A.: A Statistical Approach to Neural Network for Pattern Recognition. Wiley-Interscience, New York (2007)
6. Forrest, S., Hofmeyer, S.A., Somayaji, A., Longstaff, T.A.: A sense of self for Unix processes. In: Proceedings of the 1996 IEEE Symposium on Security and Privacy, pp. 120–128. Anonymous (1996)

7. Hayes, M.J.: Monitoring the 1996 drought using the standardized precipitation index. Bull. Am. Meteorol. Soc. **80**, 429 (1999)
8. Hughes, B.L., Saunders, M.A.: A drought climatology for Europe. Int. J. Climatol **22**, 1571–1592 (2002)
9. Jin, H., Jiang, W. (eds.): Handbook of Research on Developments and Trends in Wireless Sensor Networks: From Principle to Practice. Information Science Reference (IGI Global), Hershey (2010)
10. Kennett, R.J., Korb, K.B., Nicholson, A.E.: Seabreeze prediction using Bayesian networks. In: Cheung, D., Williams, G.J., Li, Q. (eds.) PAKDD 2001. LNCS (LNAI), vol. 2035, pp. 148–153. Springer, Heidelberg (2001)
11. Kjaerulff, U.B., Madsen, A.L.: Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis, 1st edn. Springer, New York (2007)
12. Lloyd-hughes, B., Saunders, M.A.: A drought climatology for Europe. Int. J. Climatol. **22**, 1571–1592 (2002)
13. Luther, K., Bye, R., Alpcan, T., Muller, A., Albayrak, S.: A Cooperative AIS framework for intrusion detection. In: IEEE International Conference on Communications. ICC '07, pp. 1409–1416. Anonymous (2007)
14. Masinde, M., Bagula, A.: A framework for predicting droughts in developing countries using sensor networks and mobile phones. In: Anonymous, pp. 390–393 (2010)
15. McKee, T.B., Doesken, N.J., Kleist, J.: The relationship of drought frequency and duration to time scales. In Proceedings of the 8th Conference of Applied Climatology, 17–22 January 1993
16. McKee, T.B., Doesken, N.J., Kleist, J.: Drought monitoring with multiple time scales. In: 9th AMS Conference on Applied Climatology, pp. 233–236 (1995)
17. Mishra, A.K., Desai, V.R.: Drought forecasting using feed-forward recursive neural network. Ecol. Modelling **198**, 127–138 (2006)
18. Mishra, A.K., Singh, V.P.: A review of drought concepts. J. Hydrol. **391**, 202–216 (2010)
19. Sajikumar, N.: A non-linear rainfall-runoff model using an artificial neural network. J. Hydrol. **216**, 32–35 (1999)
20. Sims, A.P.: Adopting drought indices for estimating soil moisture: a North Carolina case study. Geophys. Res. Lett. **29**, 1183 (2002)
21. Watkins, A., Timmis, J., Boggess, L.: Artificial immune recognition system (AIRS): an immune-inspired supervised learning algorithm. Genet. Program. Evol. Mach. **5**, 291–317 (2004)
22. Witten, I.H., Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann Series in Data Management Systems, 2nd edn. Morgan Kaufmann, San Francisco (2005)
23. Xu, C., Li, T., Huang, X., Jiang, Y.: A weather forecast system based on artificial immune system. In: Wang, L., Chen, K., Ong, Y.S. (eds.) ICNC 2005. LNCS, vol. 3611, pp. 800–803. Springer, Heidelberg (2005)