International Journal of Research Studies in Science, Engineering and Technology
Volume 2, Issue 10, October 2015, PP 42-46
ISSN 2349-4751 (Print) & ISSN 2349-476X (Online)

Knowledge Discovery from Time Series Data by Using Neural Network with Unsupervised Learning

Agarwala Basanthi (*Student*)
Department of Computer Science and Systems Engineering
AU College of Engineering, Andhra University
Visakhapatnam, India

Dr. A Mary Sowjanya (*Asst. Professor*)
Department of Computer Science and Systems Engineering
AU College of Engineering, Andhra University,
Visakhapatnam, India

Abstract: *Data mining massive datasets for knowledge discovery is a latest trend in business and academia in recent years. Real-world data are usually of time series, consisting of categorical and numeric attributes. Mining knowledge from massive, time series data is a challenge. To explore unknown data, visualised analysis allows users to gain some initial understanding regarding the data and to prepare for further analysis. Self organizing map (SOM) has been commonly used to generate maps to represent high order data in simple 2-dimensional plane. Interesting or hidden patterns can be discovered by visual analysis with chance of discovering valuable knowledge. In previous studies, an extended SOM has been put forward to visualise this type of data. However, the model works under the setting of supervised learning in order to measure the similarity between categorical values. The proposed model can work under unsupervised learning in the absence of class attributes or domain experts. Experimental results are reported to demonstrate effectiveness of the proposed approach.*

Keywords: *Self Organising Map, Unsupervised Learning, SOM, Time Series, Knowledge Discovery, Neural Network.*

1. Introduction

Nowadays, big data in big companies contain lot of valuable knowledge and various hidden patterns. Analysing big data to discover these hidden patterns is a latest trend. But it is not an easy task. Not only is the volume of the data huge but also real-world data usually consist of different types of attributes such as categorical and numeric attributes. Most of this data is generated over a period of time which includes measurements taken on interval basis. This generates organised data called time series. Most of algorithms handle only one type of values. When this type data are encountered, a preprocess transforming one type of the data to the other is performed prior to using the algorithms.

One of the standard methods used to transform a categorical attribute is 1-of-k coding which will generate binary attributes from categorical attributes. The corresponding binary attribute to the categorical value is set to one and the others zero. The 1-of-k coding style has a disadvantage: Significance of the categorical values is lost after conversion to binary values. For instance, the set of binary values does not reflect that a drink Coke is more similar to Pepsi than to Latte. The data is then converted into percentages, so that it is easy to compare on value with the other.

Self organising map (SOM) is a popular neural network used for visualised clustering analysis. It can project high-dimensional data to a low-dimensional space, normally a two dimensional one. So that the high-dimensional data can be analysed on the two dimensional map. Moreover, when projecting data to a low-dimensional space, SOM can preserve the topological order in the data. That is, data close to one another in the data space are also near to one another on the map. However, when mixed-type data are encountered and 1-of-k is resorted to convert the data, topological order shown on the projection map will be distorted due to the loss of semantics after the transformation. Generalised SOM model or GSOM can solve this issue. GSOM not only allows processing mixed-type data directly but also can preserve the semantics. A distance hierarchy data structure which is a tree, consisting of weights, nodes and links have been used to depict the relationship between the value with categorical values as leaf nodes. The distance between each node in a link is associated with some weight. The total weight between 2 nodes in a path is used to measure the distance. Since, in this data structure, all the leaf nodes in a subtree will be of similar type, it will be easy to maintain the semantics of the values that were lost during 1-of-k coding conversion.

The main objective of this approach is to provide a way to discover knowledge in the time series data having mixed attributes by using an unsupervised approach.

2. Background

2.1. 1-of-k Coding

The primary goal for this method is to convert categorical values to binary values. But is has disadvantage of losing the semantics in the dataset after conversion to binary values. It can be better explained with an example of a typical day weather can be hot or cold and weather might be cloudy, raining, snowing sunny.

The following tables illustrates how the collected data will look like and how it will look after conversion to binary values using this method.

Table1. *Sample Data*

ID	Temperature	Climate
1	hot	summer
2	cold	cloudy
3	cold	raining
4	cold	snowing

After conversion

Table2. *Sample Data after 1-of-k coding conversion.*

ID	hot	cold	summer	cloudy	raining	snowing
1	1	0	1	0	0	0
2	0	1	0	1	0	0
3	0	1	0	0	1	0
4	0	1	0	0	0	1

2.2. Agglomerative Hierarchical Clustering

It is a bottom-up clustering method where clusters have sub-clusters, which in turn have sub-clusters, etc. It starts with every single object in a single cluster. Then it merges the closest pair of clusters by satisfying some similarity criteria in successive iterations, until all of the data is in one cluster.

$$d(y_i,y_j) = \frac{1}{|C|}\sum_{c \,\varepsilon\, C} |f(y_i \Rightarrow c) - f(y_j \Rightarrow c)| \qquad (1)$$

where y_i and y_j are the different values in a class attribute C, $f(y_i \Rightarrow c)$denotes ratios of co-occurrence of y_i and c with respect to y_i.

As shown in the Figure 1, snowing and rainy will be more close than cloudy and summer.

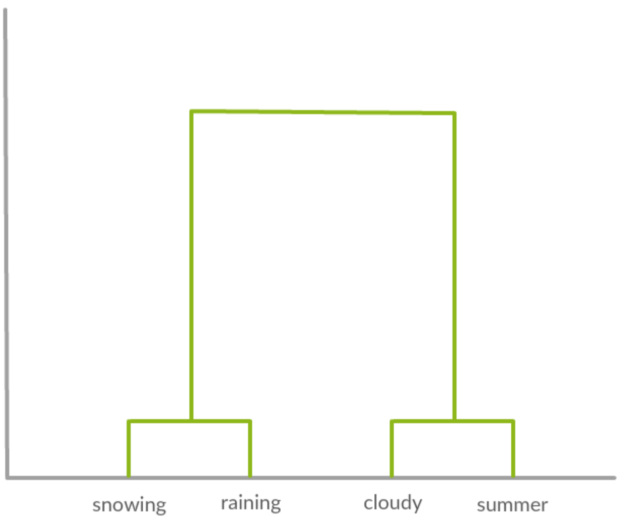

Fig1. *An example of distance hierarchy constructed from a pair wise distance matrix by using hierarchical clustering*

3. Methodology

The following process will demonstrate the implementation of unsupervised learning without having class attributes in the data set. The approach is as follows.

Assume A and B are two distinct values in a categorical attribute. A and B are deemed to be similar or have a small distance if the situation of A co-occurs with the values in the other feature attributes is very similar to the situation of B co-occurs with those values.

3.1. Distance between Categorical Values (DCV)

This unsupervised approach to calculating the distance between two categorical values in a target attribute Y uses the information of context attributes *X* of *Y*. A context attribute can be categorical or numeric. For a categorical context attribute X_c*(belongs to X)*, the distance between y_i and y_j in *Y* is dependent on the difference of conditional probabilities of y_i and y_j given x_k in X_c. For a numeric context attribute X_n, the distance is dependent on the difference between the averages of the values in X_n which co-occurs with y_i and y_j, respectively. With consideration of all context attributes, the distance between y_i and y_j is defined as in equation (2) where $Avg(X_{n,*})$ is the average which comes from the numeric values of co-occurrence of y in context attribute Xn and we use the deviation between the maximum and the minimum value in X_n to normalise the value to [0,1].

$$d(y_i, y_j) = \frac{1}{|X|}\left(\sum_{X_c \in X}\sqrt{\frac{\sum_{x_k \in X_c}\left(P(y_i|x_k) - P(y_j|x_k)\right)^2}{|X_c|}} + \sum_{X_n \in X}\frac{|Avg(X_{n,j}) - Avg(X_{n,i})|}{Max(X_n) - Min(X_n)}\right) \quad (2)$$

An Example: Two approaches, 1-of-k coding and DCV for measuring the distance between categorical values have been presented. A simple synthetic dataset, shown in Table 3, is used for the calculation of the distance. Table 1 is a toy dataset for illustration. By 1-of-k, A will be transformed to a list of two binary attributes, namely,. The value of A of the first transaction is therefore transformed to <1, 0> while that of the fourth <0, 1> accordingly.

The proposed DCV considers context attributes instead as shown in its equation. Accordingly, the distance between a and b is therefore

$$d(a,b) = \frac{1}{2}\left(\sqrt{\frac{D_c}{4}} + \frac{|D_B|}{90-20}\right) \quad (3)$$

Note that D_B and D_C are context attributes B and C in the example.

D_B = |Avg(D_B,a) - Avg(D_B,a)|

= |50 - 71.66|

= 21.66

D_C = (a/X - b/X) + (a/X - b/X) + (a/X - b/X) + (a/X - b/X)

= 0 + 0 + 1 + 1

= 2

Therefore the distance is d (a, b) = 0.51

3.2. Evaluation Measures

We use two measures, mean square error (MSE) and entropy, to evaluate the projection results of SOM with different approaches to handling categorical values. MSE measures the average distance between the input instance and the weight of its best match unit and is defined in equation (4).

$$\text{MSE} = \frac{\sum_{i=1}^{|X|} ndist(x_i, BMU_i)^2}{|X|} \quad (4)$$

where |X| is the number of input data and is an instance of X. is the prototype of the neuron into which is projected. The ndist(.,.) is the normalised distance between the two arguments. The distance is normalised by the number of attributes to allow performance comparison among different approaches. Note that 1-of-k coding increases the number of attributes.

Table3. *A Sample illustrative dataset*

ID	A	B	C	Class
1	a	50	X	Yes
2	a	20	Y	No
3	a	80	Z	No
4	b	60	P	Yes
5	b	65	Y	Yes
6	b	90	A	No

Entropy measures consistence of class labels of instances projected in neurons and is defined by the weighted average of entropies of individual neurons. It is defined in equation (5) where |Xn| is the number of data projected into neuron n.

$$\text{Entropy} = \frac{|X_n|}{|X|}\sum_{n=1}^{N} Entropy(n) \tag{5}$$

4. Experimental Results

The data set used is Ozone level detection. The data has 73 attributes but finally 15 were retained. The summary of the data set is as shown below along with the analysis of the performance of both the algorithms.

Table4. *Summery of Ozone Level Detection dataset*

Dataset	Data Points	Total Attributes	Class
Ozone level detection	2536	73	1

The projected result of the dataset by the final algorithm is shown below.

Figure.2 shows the variations of thickness of ozone layer. The other Figure.3 shows the heat-map of factors the affect the Ozone layer. This include different chemicals released in environment, wind and many others. These pictures show a clear effect of the negative factors on ozone layers by highlighting the regions on heat-maps.

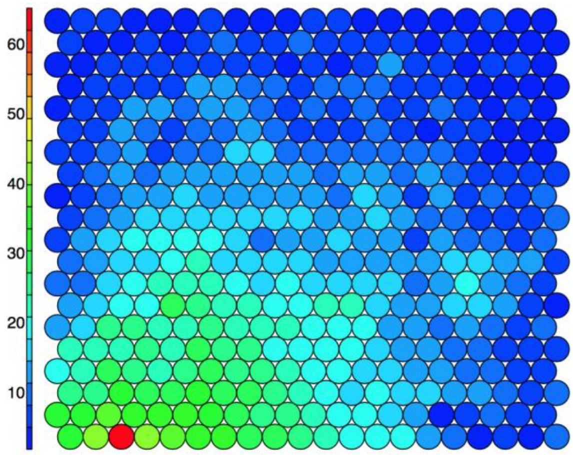

Fig2

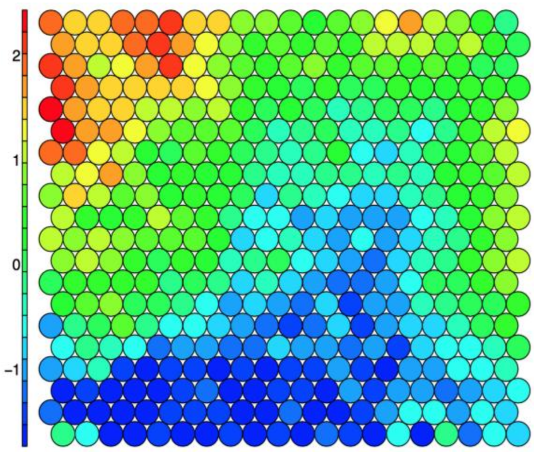

Fig3

5. Conclusion

As shown above, Self Organising Maps (SOM) can be used to generate better visualisations of trained neural network on the dataset. It has been shown that unsupervised learning on a classless dataset can be achieved by use of appropriate implementation of our algorithm. To better understand the multi-dimensional time series data, all the columns are converted to distances with respect to each other and perform the necessary complex calculations. After these calculations, data is fed to the neurons in SOM for training. Once the training is complete, it is possible to represent the multi-dimensional dataset in a two dimensional plane. The SOM visualisations can be used to understand the impact of one column on the other. This way, hidden knowledge can be discovered in a classless time series multivariate dataset using neural network and unsupervised learning.

REFERENCES

[1] Chung-Chian Hsu, Generalizing self-organizing map for categorical data, IEEE Transactions on Neural Networks, v.17 n.2, p.294-304, March 2006

[2] Esling, P. and Agon, C. 2012. Time-Series data mining. ACM Comput. Surv. 45, 1, Article 12 (November 2012).

[3] Warren Liao T (2005) Clustering of time series data—a survey. Pattern Recognit 38(11):1857–1874

[4] D. Kotsakos, G. Trajcevski, D. Gunopulos, C. Aggarwal. Time-series Data Clustering,, in "Data Clustering: Algorithms and Applications", CRC Press, 2013.

[5] Peter J. Brockwell, Richard A. Davis. Time Series: Theory and Methods Springer Series in Statistics, ISSN 0172-7397.

AUTHORS' BIOGRAPHY

Agarwala Basanthi

I received B.Tech in the year 2012 from Rajahmahendri Institute of Engineering and Technology, (Affiliated to JNTUK). Now pursuing M.Tech in College of Engineering, Andhra University, Visakhapatnam, A.P, India.

A. Mary Sowjanya

Received her B.Tech and M. Tech Degree in Computer Science & Engineering from Andhra University. Completed her PHD in CSSE from Andhra University. She specialises in Data Mining with an experience of more than 2 years. Currently working as Assistant Professor in Department of Computer Science and Engineering in Andhra University, Visakhapatnam.