Analysis of Social Media Networks based on Self-organizing Feature Maps Approach for Social E-commerce Networks

Mohamed E. A. Ebrahim
Computer Science Department
Faculty of Graduate Studies for
Statistical Research, Cairo
University, Giza, Egypt

Lamia Fattouh Ibrahim
Faculty of information systems and computer
science of October 6 University, Giza, Egypt;
Computer Science Department Faculty of
Graduate Studies for Statistical Research, Cairo
University, Giza, Egypt

Hesham A. Salman
Computer Science Department,
Higher Institute of Computers
and Information Technology,
Alshrouk Academy, Cairo, Egypt,
Dr.heshslman@sha.edu.eg

Abstract

Kohonen's algorithm is a computational data analysis method. Social E-commerce mining requires human data analysts and automated software programs to sift through massive amounts of raw social media data. Social commerce networks using user frequent pattern mining.

This paper focused on discern patterns and trends relating to social media usage based on self-organizing map SOM algorithm for analysis of social media networks and websites to get user access patterns characteristics.

Key Words: Kohonen's algorithm; Social E-commerce network; web mining; neural network

1. INTRODUCTION

The self-organizing map (SOM) has been used vastly as an engineering gadget for information observation process and data dissection, the SOM can be viewed as a clustering algorithm which produces a set of clusters organized on a ordinary network [1, 14]. The Kohonen's algorithm is a mapping of a high dimensional space to a low dimensional space. The aim of the SOM is to generate a topology preserving mapping, where the neighborhood relations in the input space are preserved as well as possible in the neighborhood relations of the units on the map [2, 12 ,16].

Customers can shop online using a range of different computers and devices, including desktop computers, laptops, tablet computers and smart phones.

A typical online store enables the customer to browse the firm's range of products and services, view photos or images of the products, along with information about the product specifications, features and prices [9].

The largest of these online retailing corporations are Alibaba, Amazon.com, and eBay provide forums for consumer ratings, evaluations and advice on users [3].

Social commerce combines e-commerce with social networking to achieve business goals and behaviors user [7].

The Information comes from social networking services and sites, which have seen spectacular prosperity in publicity in the last decade, and user-generated contents such as, blogs, wikis, discussion forums, posts, chats, tweets, podcasting, pins, digital images, video, audio files, and other forms of media created by users of an online systems. Information is most functional for product marketing and maturing, brand familiarity, and advice and help through discourse groups [4].

2. CONVOLUTIONAL NEURAL NETWORKS(CNNS)

Many types of research had been done for overcoming challenges web mining over social media networks, in order to improve performance E-commerce services. They had used different simulators to achieve their goal. In [3] the authors used experiments with veritable and synthetic data sets are considered and comparative study of the proposed networks with fuzzy c-means methods of the literature of symbolic data analysis for interval data was done based on SOM. The authors of [4] combined the cross-case synthesis technique and the negative case analysis approach to test the applicability, robustness, and the generality of the proposed system for enhance social commerce. The results of [5] illustrate that SOM is notable alternative to the main component analysis. According to the results, the classification errors were equal or less for SOM. Clustering using 1-dimensional SOM is good alternative to the conventional k-means algorithm. 1-dimensional SOM converges faster for pattern recognition process. In [12] The results used an image compression approach based on the wavelet transform and the vector quantization by Kohonen's network and the Huffman coding, to improve the reconstructed image quality and the compression ratio.

The authors have made an excellent effort to improve performance classification of patterns; we participated in the effort to reach a new research point by using analysis social media access users patterns based on Kohonen's Self-Organizing Feature Map SOM in two scenarios to reach the best results. This work presents a simulation to applying Neural Network based Clustering using SOM.

3. KOHONEN SELF-ORGANIZING FEATURE MAP (SOM)

Kohonen's Self-Organizing Feature Map SOM is a neural network which modifies itself as a reaction to the input patterns. This feature is called self-organization and it is perform using competitive learning [2]. The basic competitive learning means that a competition process takes place before each cycle of learning. In the competition process a winning processing element is chosen by some standards. Often this standard is to minimize a Euclidean distance between the input vector and the weight vector [13 ,15].

After the winning processing element is selected, its weight vector is acclimatized according to the learning law used. SOM differs from the basic competitive learning so that rather than adapting only the winning processing element, also the neighbors of the winning processing element are acclimatized. The self-organization feature of SOM is based on the use of the neighborhood of the winning processing element [5].

SOM establishes topologically ordered mappings between input data and processing elements of the map. Topological order means that if two inputs are similar, then the most active processing elements responding to the inputs are situated near each other in the map and the weight vectors of the processing elements are coordinated to an ascending or descending order [6, 8].

This version of SOM learning algorithm is based on:

1. **Initialization-** Set weights to small random values w_j.

2. **Sampling-** Select input randomly **x** from the dataset.

3. Calculate Euclidean distance to all processing elements [13].

$$d_j = \sqrt{\sum_{i=1}^{m} (x_i - W_{ij}(t))^2} \ (j = 1, 2, \ldots, n)$$

4. **Matching-** Choose the winning processing element $I(\mathbf{x})$ best matching unit with the minimum distance of

$$d_j(\mathbf{x}) = \sum_{i=1}^{D} (x_i - w_{ji})^2$$

5. **Updating-** Update weight vectors of the winning processing element.

$$\Delta w_{ji} = \eta(t) \ T_{j,I(\mathbf{x})}(t) \ (x_i - w_{ji})$$ Where $T_{j,I(\mathbf{x})}(t)$ is a Gaussian Neighborhood and $\eta(t)$ is the learning rate.
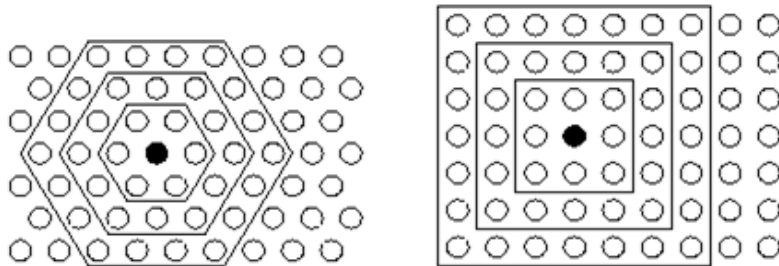

Figure 1. (a) Hexagonal grid   (b) Rectangular grid

The gain term should be near unity at the start of the practice and reduce during practice. The exact rule for reducing is not essential; it can be some linear or nonlinear function of time. First, the size of the neighborhood (time1) is large making hard organizing of SOM possible and the size is reduced (time2) and (time3) as time, shown on Figure 1. Finally, in the end only the winning processing element is acclimatized making the accurate tuning of SOM conceivable. The use of neighborhood makes topological arranging process conceivable and jointly with competitive learning makes the process nonlinear [5, 6].

6*. Continuation-* Go to step 2 or stop the repetition when adequate inputs are presented.

4. IMPLEMENTATION AND RESULTS

According to Kohonen's neural networks algorithm, the users of social media accessing model was performed, which involves two steps [3, 11]:

The first step of the model composed of a feed neural network which feeds to a Kohonen's Self-Organizing Map SOM in the second step of the model. Feed neural network is used to learn and foretell the scores of types from social media websites.

The Kohonen's algorithm incorporates the models with the learning to predict and update strategies of the general Kohonen's algorithm. Since Kohonen's algorithm is an optimal procedure for data analysis and information monitoring process of the social media website visits [3, 7].

The SOM is a square grid of n-square neurons, sorted in n rows and n columns and assign n. n should be between 2 and 10. The cycle consists of submitting all the notices to the map and specifying the number of cycles.

In each cycle all the notices are presented to the map once. Order in which they are presented could be random - hence varies from cycle to cycle or could be as the original order in which it had entered in the Data sheet, the end result of the learning parameter is less than the start value and both the values are robustly between 0 and 1, the end value of Sigma is less than the start value and both values are robustly between 0% and 100%.
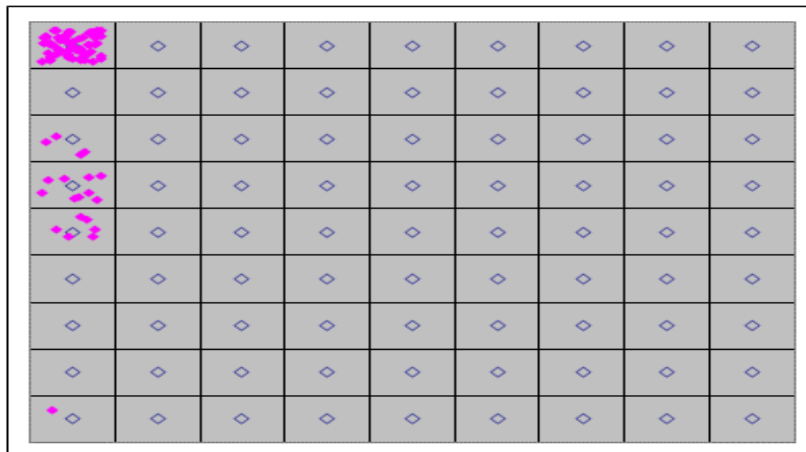
As the training of the map progress, both Learning Parameter and Sigma are decreased from the initial value to the end value and select the rate of this decay as linear or exponential

While training the map real variables in the data are scaled so that for each variable - values are between -1 and 1. This is called Normalization of data. Normalization takes a long time - especially for big data sets as shown in Table 1.

The map body composed of 9×9 grid, which means a neural network with $M = 81$ neurons as shown in Figure 2. The goodness of the performance of the $n = 100$ multidimensional data points by the neuronal vectors wj is weighted by two indices: the quantization error and the topological error. For which the first- and second-best matching vectors are not adjacent neighbors in the map.

**Table 1. Kohonen's neural network**

Neural Network based Clustering			
(Using Kohonen's **S**elf **O**rganizing **M**aps (**SOM**))			
		Learning parameter *(should be >0 and <1)*	
Number of Observations	**80**	Initial value	**0.9**
(Needs to be between **5** and **5,000**)		Ending value	**0.1**
		Decay	**Exponential**
Number of Variables	**9**		
(Needs to be between **3** and **50**)		Sigma for the Gaussian neighborhood	
		as % of map width *(should be > 0% and < 100%)*	
Enter **n**, where **n**-Square = # neurons in the map	**9**	Initial value	**50.0%**
(n needs to be between **2** and **10**)		End value	**1.0%**
		Decay	**Exponential**
[Note: By entering **n** you are specifying that the maximum number of clusters will be at most **n**-square. e.g. if you enter **n** = 4, you will get less than or equal to 16 clusters]			
Number of training cycles	**100**		
(Needs to be between **1** and **1000**)			

Figure 2. Output weights with five clusters

There are two scenarios done in this work with different training data sets and neurons in the map equal nine.
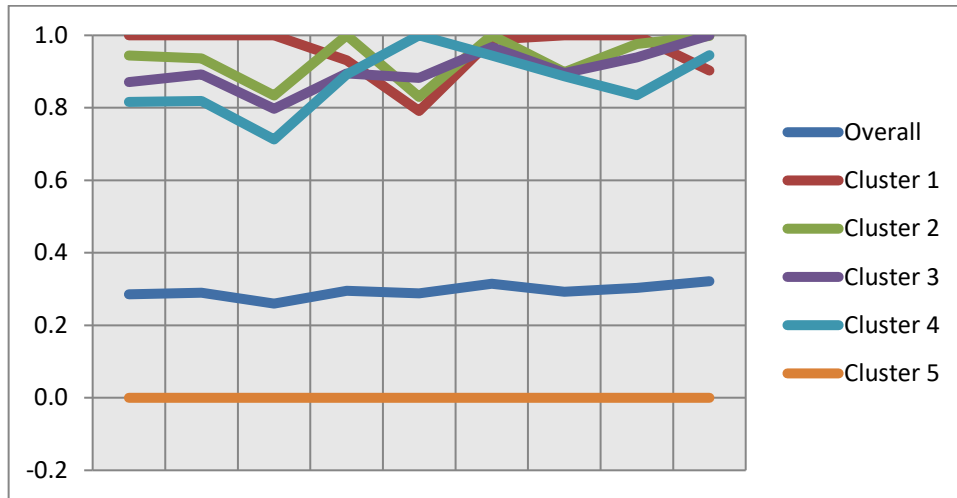
Figure 3. Output clusters

First scenario is Training data set with nine variable input and neurons in the map equal nine, the result is five clusters as shown in figure 3 and table 2.

Table 2. Shows template data result with number of variables equal nine

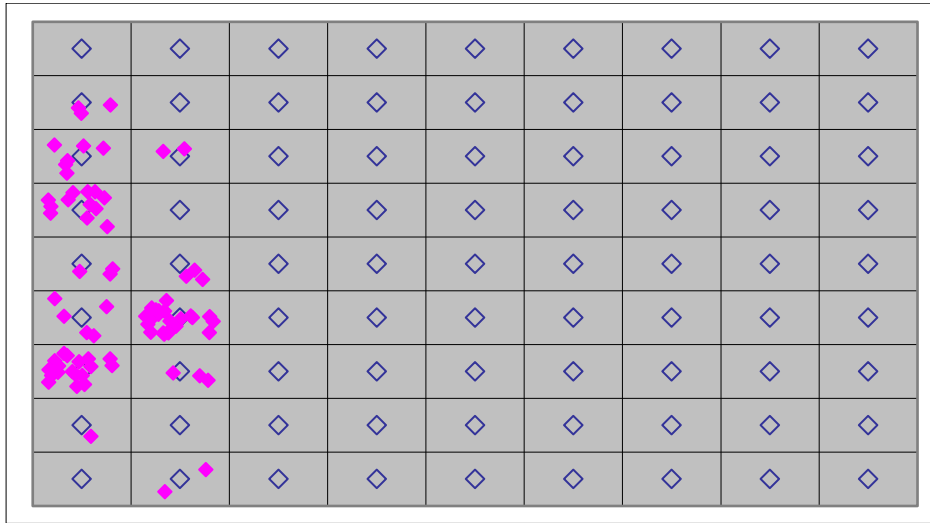| Variables	Overall	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Y1	0.3	0.8	0.9	1.0	1.0	0.9
Y2	0.3	1.0	0.9	1.0	0.9	0.9
Y3	0.3	1.0	0.9	1.0	0.9	0.9
Y4	0.3	0.9	1.0	1.0	1.0	0.9
Y5	0.3	1.0	0.9	0.8	0.9	0.9
Y6	0.3	1.0	1.0	1.0	1.0	0.9
Y7	0.3	1.0	0.9	0.9	0.8	0.8
Y8	0.3	1.0	1.0	1.0	1.0	0.9
Y9	0.3	0.9	1.0	1.0	1.0	0.9

Figure 4 Output weights with twelve clusters

Table 3. Shows template data result with number of variables equal three

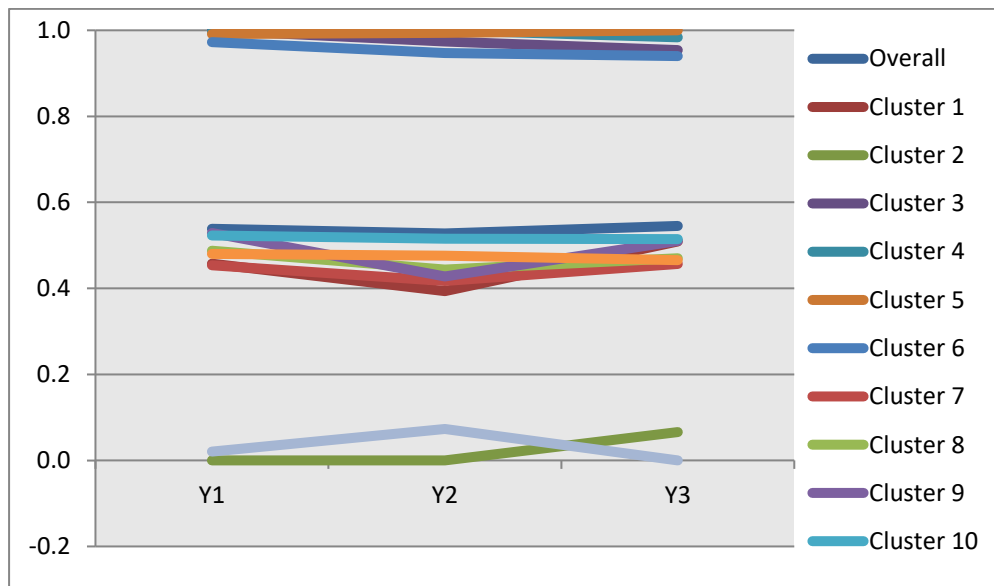| Variable	Overall	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10	Cluster 11	Cluster 12
Y1	0.5	0.5	0.0	1.0	1.0	1.0	1.0	0.5	0.5	0.5	0.5	0.5	0.0
Y2	0.5	0.4	0.0	1.0	1.0	1.0	0.9	0.4	0.4	0.4	0.5	0.5	0.1
Y3	0.5	0.5	0.1	1.0	1.0	1.0	0.9	0.5	0.5	0.5	0.5	0.5	0.0

Figure 5. Output clusters

Second scenario is Training data set with three variables input and neurons in the map equal nine as shown in figure 4, the result is twelve clusters as shown in figure 5 and table-3.

5. CONCLUSION

This paper has introduced a proposed methodology to improve E-Commerce networks.
By social media networks visits to analysis Social media data mining based self-organizing map SOM. This paper used number of different variables of social media networks by Kohonen's algorithm neural network. The result is number of clusters expresses users orientations. Future work will be making a comparison between others algorithms like Principal component analysis PCA and Self-organizing map SOM for analysis Social media data mining for example.

REFERENCES

[1] Samuel Kaski. "Self-Organizing Maps". Helsinki University of Technology, Helsinki, Finland. 2017.
[2] J. Lampinen and E. Oja." Fast Computation of Kohonen Self-Organization" Lappeenranta University of Technology, Finland 1990.
[3] DaweiLiu, ZilongZhang, XiaohongGuo. "Web mining based on one-dimensional Kohonen's algorithm: analysis of social media websites". 2017.
[4] Hui Hanand, SilvanaTrimi. "Social Commerce Design: A Framework and Application" Journal of Theoretical and Applied Electronic Commerce Research VOL 12 / ISSUE 3 / SEPTEMBER 2017.
[5] Markus Törmä. "KOHONEN SELF-ORGANIZING FEATURE MAP IN PATTERN RECOGNITION". Photogrammetric Journal of Finland,VOL15/ ISSUE 1/Published – 1996.
[6] Jari Kangas. "On the Analysis of Pattern Sequences by Self-Organizing Maps". Thesis for the degree of Doctor of Technology, Helsinki University of Technology, Laboratory of Computer and Information Science, Espoo, Finland.1994.
[7] Ahmed al-shami "Unified knowledge-based economy hybrid forecasting". Technological Forecasting and Social Change, 91, pp.107-123. 2015.
[8] Pankaj Kumar Kashyap, Kirshna Kumar, and Sushil Kumar. "Fuzzy-Kohonen Self-organizing Clustering Algorithm in Wireless Sensor Networks", 2018.
[9] Trafialek J, Laskowski W, Kolanowski W. "The use of Kohonen's artificial neural networks for analyzing the results of HACCP system declarative survey", 2015.
[10] Bourgeois N, Cottrell M, Deruelle B, Lamasse S, Letremy P. "How to improve robustness in Kohonen maps and display additional information in factorial analysis: application to text mining. Neuro computing". 2015.
[11] Anna M. Bartkowiak and RadosławZimroz. "Complementary View on Multivariate Data Structure Based on Kohonen's SOM, Parallel Coordinates and t-SNE Methods", 2018.
[12] Mourad Rahali, Habiba Loukil, and Mohamed Salim Bouhlel. "Analysis of Image Compression Approaches Using Wavelet Transform and Kohonen's Network" 2018.
[13] EwaSkubalska-Rafaj_lowicz. "Exploring the Solution Space of the Euclidean Traveling Salesman Problem Using a Kohonen SOM Neural Network". 2017.

[14] Lyes Khacef, L aurent R, Benoit M (2020) "Improving Self – Organizing Maps with Unsupervised Feature Extraction" International Conference on Neural Information Processing, ICONIP 2020 PP474-486.

 [15] Wang L, Chen Y, Jiang H, Yao J (2020) Imbalanced credit risk evaluation based on multiple sampling, multiple kernel fuzzy self-organizing map and local accuracy ensemble. Applied Soft Computing 91: 106262. https://doi.org/10.1016/j.asoc.2020.106262

 [16] Murugesan, V.P., Murugesan, P. Some measures to impact on the performance of Kohonen self-organizing map. Multimed Tools Appl 80, 26381–26409 (2021). https://doi.org/10.1007/s11042-021-10912-1