Neural Comput & Applic
DOI 10.1007/s00521-016-2410-9

ORIGINAL ARTICLE

Web mining based on one-dimensional Kohonen's algorithm: analysis of social media websites

Dawei Liu[1] · Zilong Zhang[1] · Xiaohong Guo[1]

Received: 17 February 2016 / Accepted: 30 May 2016

Abstract One-dimensional Kohonen's algorithm is a process of mining knowledge which finds the characteristics of social media websites as a mode from the sequence database. Social media web log records generated constantly, and user access patterns will change accordingly. This study focused on taking advantage of the dynamic characteristics of the Kohonen algorithm, delivering a fast and efficient incremental mining algorithm and testing the new developed model.

Keywords Kohonen's algorithm · Social media · Data mining · Neural computing

1 Introduction

In this work, a variant to self-organizing maps based on Kohonen's algorithm is proposed in order to analyze a real B2C websites. The Kohonen algorithm is a neural network proposed by Kohonen, which maps a distribution of vectors of any dimension into a lower-dimensional space, generally one or two, while maintaining a high degree of topological ordering, or neighborhood preservation. It is widely adopted to visualize high-dimensional data sets and to mine online data [1–4].

Social networks such as eBay.com, Amazon.com and Taobao.com (the largest online shopping platform in China) provide forums for consumer ratings, evaluations and advice on users. Consumers using Taobao.com spent USD 180 billion in 2014, and 94 % of their purchases were shared with others [5]. Although social commerce has become an important topic for many researchers, previous studies of social commerce have generally been developed from the literature of e-commerce. Social commerce is online business that combines e-commerce with social media (e.g., twitter) and social networking to accomplish business goals, functions and behaviors. Qu defined social commerce as online business activities initiated via social media which entails business transactions through either social media (e.g., Taobao.com) or other e-commerce sites. Social commerce enables the use of various social technologies to improve the shopping experience for customers. Smart phones are now the most commonly used tools in social commerce transactions.

Social commerce adoption is regarded by the computer and information industries as the 'future generation' of digital business because of their convenience in social interaction and Internet login [6]. Since smart phones and tablet computers are widely adopted in social interaction and electronic commerce, wearable devices, as the extended or future generation of digital devices, should provide better experiences of social interaction. Therefore, understanding the factors impacting the access of social websites can lead to better academic integrity and market applications.

One of the main drawbacks associated with the classical Kohonen's algorithm results from the prerequisite the adequate size of the output layer in advance [7, 8]. In this work, we extended the existing Kohonen's model, which allows shortening the time of training with regard to other approaches that also try to solve the problem of the optimal size of SOM, e.g., the growing hierarchical self-organizing maps algorithm [9].

✉ Dawei Liu
hduldw@163.com

[1] School of Management, Hangzhou Dianzi University, Hangzhou, China

Published online: 06 June 2016

2 Construction of a one-dimensional Kohonen's algorithm

Classic Kohonen's algorithm consists of a set of neurons usually arranged in a one- or two-dimensional grid [1]. Although higher-dimensional grids are also possible, they are hardly ever used because of their problematic visualization [8, 10, 11]. Every neuron has a fixed position in the grid and is represented by an n-dimensional weight vector, as $w = [w_1, w_2, \ldots, w_n]$.

Where n is the dimensionality of the input space. A user pattern x is randomly chosen from the data set on each training step. Then, the neuron whose weight vector is the most similar to the user pattern is searched, being this neuron the so-called best matching unit (BMU) [12]. The weight vectors of the BMU and its neighborhood are updated as follows.

Where t stands for the iteration number, $a(t)$ is the learning rate, and $h(t)$ is the neighborhood kernel, whose center is located at the BMU. The neighborhood kernel determines which neurons around the BMU are updated, and how this update functions each neuron [9].

$$w^{t+1} = w^t + \alpha(t)h(t)(x - w^t) \quad (1)$$

If the input level (first level) contains m neurons, which indicates the dimensions of the input vectors are m, the output level (second level) has n neurons [13]. Therefore, the weight coefficient number of input level and output level is $m \times n$. The input vector is $X = [X_1(t), X_2(t), \ldots, X_m(t)]^T$, where $X_i(t)$ is the component of dimension i of sample at time t; The output vector is $Y = [Y_1(t), Y_2(t), \ldots, Y_n(t)]^T$, and $W_{ij}(t)(i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, n)$ is the weight coefficient from input knot i to output knot j at time t.

Kohonen's neural algorithm follows competitive learning strategy [14]. For each input sample vector, the coefficient of weight of the winner and its relevant neurons in competitive are modified and the rest ones remain the same. The algorithm of the competition is:

Step 1 Neural network initialized
When $t = 0$, $W_{ij}(0)$ is supposed to be a small random number to set up a initialized neighborhood around knot j and its radius is $N_j(0)$;

Step 2 A vector is input as a new sample:
The new vector is $X = [X_1(t), X_2(t), \ldots, X_\mathrm{m}(t)]^T$, where $X_i(t)$ is the component of dimension i of sample at time t. The Euclidian distance of the input sample and each neuron of output levels is d_j, which is:

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

Step 3 The optimal matching output neuron
The output neuron is the one with minimum distance to input X, which is c. W_c means the coefficient vector of chosen neuron and shown as:

$$\|X - W_c\| = \min\{d_j\} \quad (3)$$

Step 4 Modify the coefficient of weight
Modifying the coefficient of weight of knot c and its neighborhood, the modified coefficient of weight should be:

$$W_{ij}(t+1) = W_{ij}(t) + \eta(t)[X_i - W_{ij}(t)] \quad (4)$$

where $\eta(t)$ is the gain item and decreases with the time to 0, and $0 < \eta(t) < 1$ would be $\eta(t) = 1/t$.

Step 5 To Step 2 until the sample ends or the gain item turns to be 0.

3 Training of Kohonen's algorithm

Triangle subjection function is adopted as the algorithm of optimization matching of output neuron for Kohonen's neural network option [15–17]. As the Kohonen neural algorithm, extended Kohonen's algorithm is a two-level feed-forward neural network with input and output levels. The input level has m neurons, and the number of neurons equals to the number of dimensions of input vectors. The output level is a one-dimensional linear matrix competitive output neuron, ranging from 0 to 1 (Fig. 1).

where $x_1, x_2, \ldots\ x_m$ are the m-dimensional input of Kohonen's neural network, marked as $X = [x_1, x_2, \ldots, x_m]^T$; $w_{11}, w_{12}, \ldots\ w_{mn}$ are the weight with $m \times n$ coefficient of input and output neurons, marked as $W_j = [w_{1j}, w_{2j}, \ldots, w_{mj}]^T$, while $j = 1, 2, \ldots, n$ and $1, 2, \ldots, n$ are the output neurons; $\mu_1, \mu_1, \ldots\ \mu_n$ are the fuzzy output values.

4 Implementation and performance results

Based on improved Kohonen's neural networks algorithm, social media users accessing model was implemented [18], which includes two stages: learning stage and application stage (shown in Fig. 2).

The one-dimensional Kohonen algorithm integrates the models with the learning rate and updating strategies of the general Kohonen algorithm. Since one-dimensional Kohonen's algorithm is a optimal procedure, the integration of one-dimensional Kohonen's models in one way solves several problems of the social media website visits (Fig. 3).

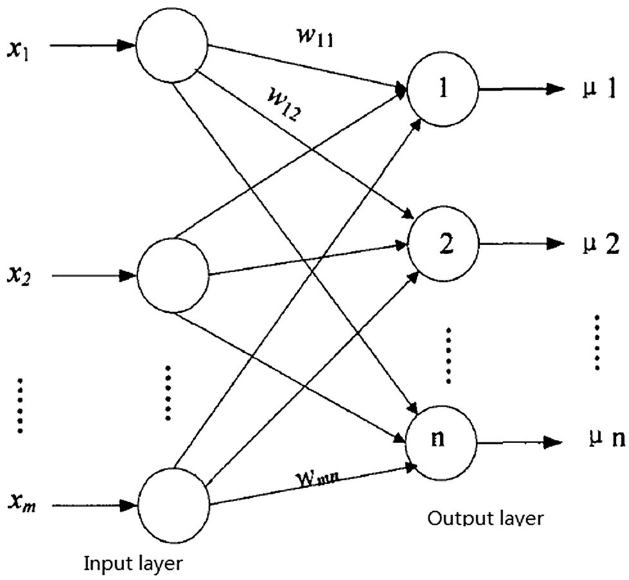

Fig. 1 Improved Kohonen's neural network

Figure 4 shows that during the evolution of Kohonen's network the initialization of the models as random neurons was originally used only to demonstrate the capability of the one-dimensional Kohonen algorithm to become ordered or organized, starting from an arbitrary initial state [19–21]. While one expects to achieve the final ordering or stable as fast as possible, so the selection of a good initial state of neuron may speed up the convergence of the Kohonen algorithms by orders of magnitude (Figs. 5, 6).

The distribution of learning rates and the size of neighborhoods are controlled by changing the value of weight exponents with time, besides the fuzzy Kohonen clustering network. Instead of decaying the learning rates and the size of neighborhood neurons to lower level or zero, it decreases the value of the weight neighborhood from a certain positive constant larger than one. Since the value of the weight neighborhood is one or two nodes, only the winner is updated with the membership value one as in the Kohonen algorithm. The improved one-dimensional Kohonen's algorithm guarantees cluster weights to converge by minimizing an objective function as in the original Kohonen model. One-dimensional Kohonen's algorithm is parallel and is independent of the feeding sequence (Fig. 7).

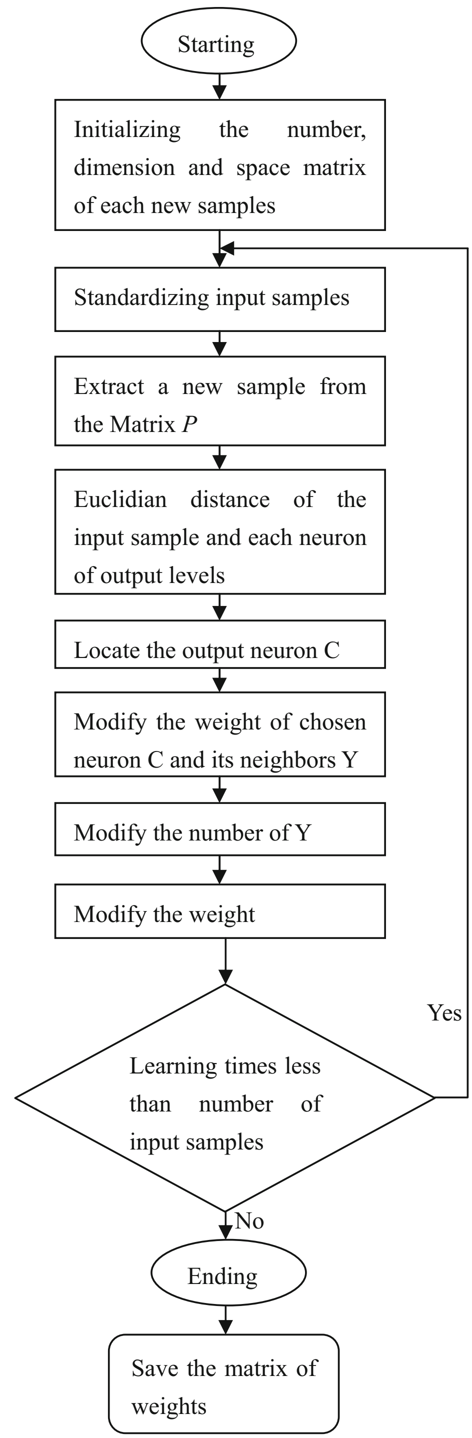

Fig. 2 Process of improved Kohonen's neural algorithms

5 Conclusion

This paper has provided insights into the different accessing patterns of social media websites by improved Kohonen's one-dimensional neural networks. The study confirms that better algorithms on social commerce users' mining could enhance their acceptance patterns on social media different from previously literature. Current research adds new knowledge regarding time and neuron matrix in existing Kohonen's models. The research helps practitioners and researchers better understand the different accessing characteristics between social media providers

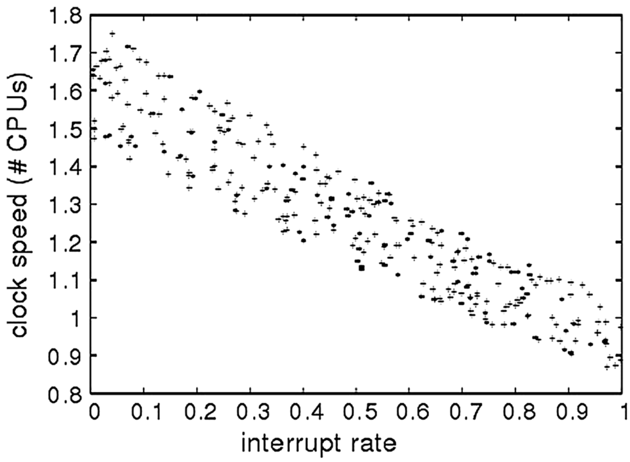

Fig. 3 Interruption rate of Kohonen's network

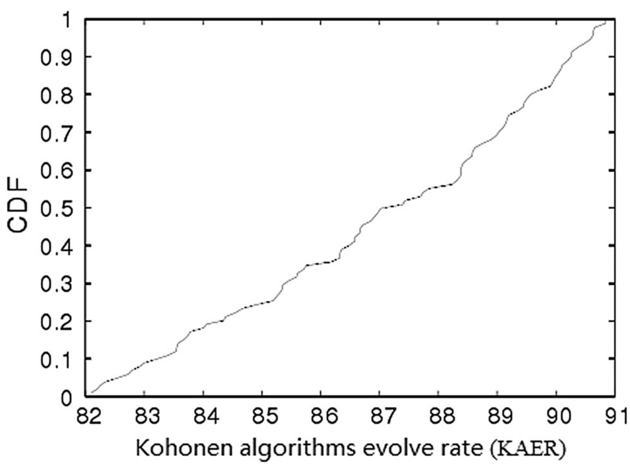

Fig. 4 Evolution rate of Kohonen's network

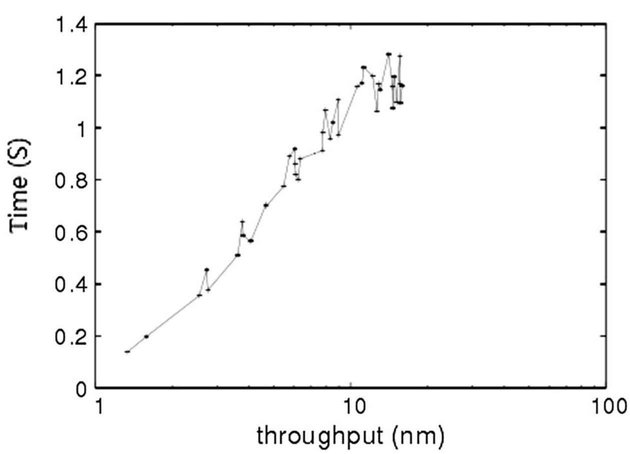

Fig. 5 Evolution results by time

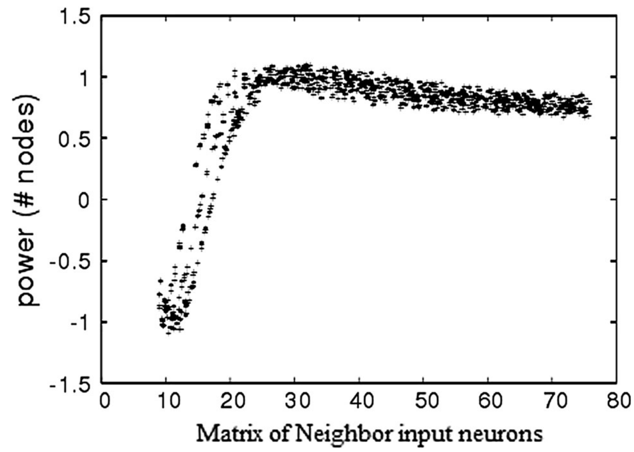

Fig. 6 Matrix of input neurons

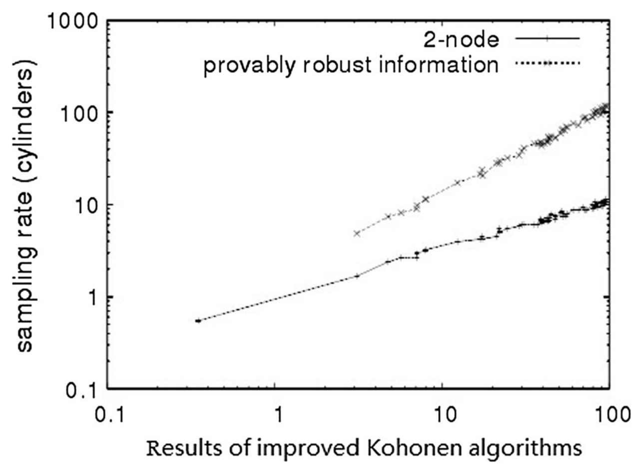

Fig. 7 Results of improved algorithms

and users. Experiments with real and synthetic data sets are considered. A comparative study of the proposed networks with fuzzy *c*-means methods of the literature of symbolic data analysis for interval data was performed. The comparison was based on an external index, the overall error rate of classification and the number of iterations needed. For the synthetic data sets, these measures were estimated by the Monte Carlo simulation method. Continued research, development and evaluation are required to provide further understanding about other potential factors that may have an impact on the acceptance of social media services in colleges and to provide useful guidelines for marketers and product designers. The results pointed out that networks introduced in this paper outperformed the methods for these synthetic and real interval data sets regarding these clustering quality measures used.

Acknowledgments The research on which this paper reports has been financially supported by Zhejiang Province Natural Science Foundation (LY15G020021) and Key Research Base of Humanities and Social Science of Zhejiang Province (ZD02-201601).

References

1. Kohonen T (1988) An introduction to neural computing. Neural Netw 1:3–16
2. Ribeiro FAL, Rosario FF, Bezerra MCM, Wagner RDC, Bastos ALM, Melo VLA, Poppi RJ (2014) Evaluation of chemical composition of waters associated with petroleum production using Kohonen neural networks. Fuel 117:381–390
3. Trafialek J, Laskowski W, Kolanowski W (2015) The use of Kohonen's artificial neural networks for analyzing the results of HACCP system declarative survey. Food Control 51:263–269
4. Bourgeois N, Cottrell M, Deruelle B, Lamasse S, Letremy P (2015) How to improve robustness in Kohonen maps and display additional information in factorial analysis: application to text mining. Neurocomputing 147:120–135
5. Lu B, Fan W, Zhou M (2016) Social presence, trust, and social commerce purchase intention: an empirical research. Comput Hum Behav 56:225–237
6. Hew J-J, Lee V-H, Ooi K-B, Lin B (2016) Mobile social commerce: the booster for brand loyalty? Comput Hum Behav 59:142–154
7. Ambroise C, Sèze G, Badran F, Thiria S (2000) Hierarchical clustering of self-organizing maps for cloud classification. Neurocomputing 30:47–52
8. Frosini L, Petrecca G (2001) Neural networks for load torque monitoring of an induction motor. Appl Soft Comput 1:215–223
9. Soriano-Asensi A, Martín-Guerrero JD, Soria-Olivas E, Palomares A, Magdalena-Benedito R, Serrano-López AJ (2008) Web mining based on growing hierarchical self-organizing maps: analysis of a real citizen web portal. Expert Syst Appl 34:2988–2994
10. Gielen C (1993) Neural computation and self-organizing maps, an introduction: by Helge Ritter, Thomas Martinetz and Klaus Schulten. Addison-Wesley, ISBN 0-201-55443-7 (hbk), ISBN 0-201-55442-9 (pbk). Neurocomputing 5:243–244
11. Cervera E, del Pobil AP (1997) Multiple self-organizing maps: a hybrid learning scheme. Neurocomputing 16:309–318
12. Kirk JS, Zurada JM (2004) Topography-enhanced BMU search in self-organizing maps. In: Yin FL, Wang J, Guo CG (eds) Advances in neural networks—Isnn 2004, Pt 2, pp 695–700
13. de Lázaro JMB, Moreno AP, Santiago OL, da Silva Neto AJ (2015) Optimizing kernel methods to reduce dimensionality in fault diagnosis of industrial systems. Comput Ind Eng 87:140–149
14. Wu J-S, Zheng W-S, Lai J-H (2015) Approximate kernel competitive learning. Neural Netw 63:117–132
15. Ruwisch D, Bode M, Purwins H-G (1993) Parallel hardware implementation of Kohonen's algorithm with an active medium. Neural Netw 6:1147–1157
16. Adeloye AJ, Rustum R, Kariyama ID (2011) Kohonen self-organizing map estimator for the reference crop evapotranspiration. Water Resour Res. doi:10.1029/2011WR010690
17. Tselentis GA, Paraskevopoulos P (2011) On the use of Kohonen neural networks for site effects assessment by means of H/V weak-motion spectral ratio: application in Rio-Antirrio (Greece). Bull Seismol Soc Am 101:579–595
18. Al Shami A, Lotfi A, Coleman S, Dostal P (2015) Unified knowledge based economy hybrid forecasting. Technol Forecast Soc Change 91:107–123
19. Kim CI, Yu IK, Song YH (2002) Kohonen neural network and wavelet transform based approach to short-term load forecasting. Electr Power Syst Res 63:169–176
20. Sanger TD (1989) Optimal unsupervised learning in a single-layer linear feedforward neural network. Neural Netw 2:459–473
21. Waller NG, Kaiser HA, Illian JB, Manry M (1998) A comparison of the classification capabilities of the 1-dimensional kohonen neural network with two pratitioning and three hierarchical cluster analysis algorithms. Psychometrika 63:5–22