International Journal of Aquatic Science
ISSN: 2008-8019
Vol 12, Issue 01, 2021

Machine Learning-Based Simulation In Remote Sensing Contexts

NurAmalina Faisal[1], Hashibah Hamid[2], MohdNorazmi Nordin[3]

[1]*Fakulti Pendidikan, UniversitiKebangsaan Malaysia*
[2]*Universiti Utara Malaysia, Sintok, Kedah, Malaysia*
[3]*Cluster of Education and Social Sciences, Open University Malaysia*

Abstract: The current study involved two proposed algorithms: K_WIC and K_CIO. The algorithm K_WIC generates the good initialization center set and supports for speeding up execution of remote sensing image clustering. The algorithm K_CIO is used to cluster remote sensing images with adding context information of each pixel. The test results show that the algorithms K_WIC, K_CIO and K_WIC_CIO (a combination of K_WIC and K_CIO) can be used effectively for remote sensing image clustering. Besides, due to the nature of the Wavelet transform, the value domain of the output data is changed. Specifically, image doesn't belonging to the domain [0,255]. Therefore, we proposed an improvement of Wavelet transformation to still ensure that the domain of the output data belongs to the domain [0,255], suitable for image data. In future work, we will continue to study the new context information and the new algorithms.

1. INTRODUCTION

Clustering is a process used to extract the main features of background objects by defining corresponding regions [1]. The task of image segmentation is from the initial multi-spectral image, proceeding to gather pixels with the same properties (color, shape, texture) into the same cluster to divide into regions and clusters. Currently, there are many different partitioning methods such as: Morphological methods, KMeans methods, Limited Gaussian mixing model (FGMM), Separation and integration, Markov models. Most the methods only use pixel intensity characteristics to perform clustering. At present, some algorithms use more contextual information in the process to reduce the complexity of segments [2]. In [3], the authors also combined fuzzy clustering algorithms and other gray level adjustment expressions to enhance medical imaging contrast. In [4], the authors used the local approach to enhance the contrast of remote sensing images. In [5], the authors improved the KMeans algorithm to use the cluster center instead.
Fuzzy CMeans algorithm is highly appreciated for image processing with the application of fuzzy clustering. It is very important that Fuzzy CMeans allows control over the number of clusters. However, the execution speed of this algorithm is very slow. With the large images like remote sensing images, the speed is even slower. In addition, the membership matrix is a major obstacle for this algorithm to perform with large images like remote sensing images.In addition, according to [6], KMeans loses the contextual characteristics (neighboring information) of each pixel when only intensity characteristics are considered. In [7], the authors proposed the 2D-KMeans algorithm with the addition of median values like spatial parameters (local context information) to increase clustering efficiency.In this study, we propose two new clustering algorithms based on KMeans. Firstly, the algrithm K_WIC combines KMeans algorithm with Wavelet technique for effective center initialization in

order to speed up the partitioning of remote sensing images. Second, the algorithm K_CIO with contextual information uses a high pass filter to enhance cluster quality.
According to [4], remote sensing is science branch remote gathering the information on the Earth surface, including sensing and taking energy released, processing, analyzing data and applying the information after analysis. Besides, most receiving systems and remote sensing images processing include seven-step process.Remote sensing images have features: image channel, space resolution, spectrum resolution, radiant resolution and time resolution. One of concrete applications is classifying land cover and building maps on the special subject in each field.

2. METHODOLOGY

K-means algorithm [7] which includes 4 steps is used to solve the K-means clustering problem and it works as follows:

Input: n objects x_i with $i = 1..n$ and clustering number c Output: clusters C_j ($j = 1..c$) to objective function E following is minimal: $E = \sum_{j=1}^{c} \sum_{x \in C_j} d^2(x, C_j)$ (1)
Step 1: Initialize the center of the clusters Select k objects C_j (j=1..c) are started center of k clusters (random or experience) *Step 2:* Attribute the closest cluster to each data point Calculate the distance between each object x_i ($i = 1..n$) and each center C_j với $j = 1..c$. The Object belongs to cluster C_S if the distance between center C_S and this object is minimal. $d(x, C_S) = \min d(x, C_j)$, $j = 1..c$ (2) *Step 3: Update centers*
Update center C_j ($j = 1..c$) by calculating the mean of all data points which belongs to that cluster. $C_j = \frac{\sum_{x \in cluster(j)} x}{count(cluster(j))}$ (3) *Step 4:* Repeat steps 2-3 until convergence

In this case, $(x,)$: the distance between object x and center Cj
In [3], authors propose the cluster center initialization algorithm CCEA (Cluster Center Estimation Algorithm) to speed up the convergence of algorithm of CDFKM.

Step	Task
1	Randomly select M subsets SB_l (l = 1 to M) of size fN from the data set S such that $SB_i \cap SB_j = \emptyset$ ($i \neq j$), where $f < 1$. Set $SU = SB_1$ and $p = 0$.
2	A set of initial cluster centers $SC_p = \{C_j\}$ and the data set SU, use CDFKM to determine a set of cluster centers $SC_{p+1} = \{C_j\}$
3	Update $p = p + 1$ and set $SU = SU \cup SB_{p+1}$. If $p \leq M$, go to step 2
4	Output SC_M as the set of initial cluster centers

In step 2, if replacing CDFKM by KMeans, we get CCEA_KMeans.

C. The Algorithm 2D-Kmeans

In [4], authors propose 2D-Kmeans algorithm. The difference between K-means and 2D-KMeans is:

With K-means, each object *x* is a vector whose components are the intensitys of channels of corresponding pixel object: *xINT*. Therefore, each center is an average vector of intensities elonging to corresponding center: *CINT*.

With 2D-KMeans, each object *x* includes kinds of components: the first components are intensities of *xINT*and the second components are intensities of local median vector of corresponding pixel object *xMED*(formula 4). Therefore, each center *C* includes 2 component kinds: the first components are average intensities *CINT* and và the second components are average intensities of median vectors: *CMED* (formula 5).

Although the K-means algorithms cluster remote sensing images effectively, the execution speed for large images like remote sensing images is still slow.

To overcome this limitation (1), we improve K-means algoritnm by the cluster centers initialization using Wavelet technique. We call this improving is algorithm K_WIC (K-means_WaveletInited Centers). Wavelet transformation is used to reduce image size. The minimal image of Wavelet transformation is used to cluster and generate the initialization center set. The chart of this algorithm is illustrated in Figure 1.

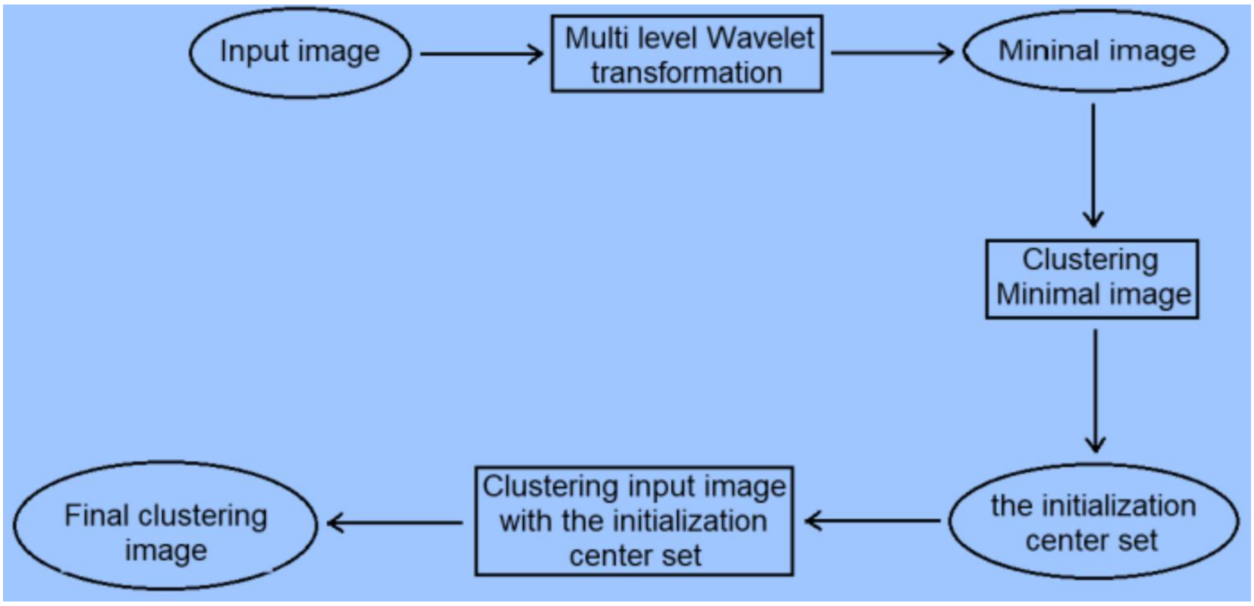

Figure 1: Wavelet transformation

3. RESULTS AND DISCUSSION

The executing time of KMeans algorithm is huge compared to the proposed algorithm K_WIC.

The executing time of the proposed algorithm K_WIC is less than the algorithm CCEAKMeans.

The second, about clustering quality:

The homogeneity of the clusters generated by the proposed algorithm K_CIO is higher than the algorithm 2D-KMeans. Therefore, the proposed algorithm considers only intensity-based features and does not take context features of pixels. This leads to the fact that the context information of pixel is lost during the process. In this paper, we present the improvements on KMeans that we call algorithms K_WIC and K_CIO to decrease computational time and maintain good clustering.

Another component comes in the form of data loss prevention (DLP). Indeed, this component is responsible for ensuring that in a given corporation, end-users do not share critical or sensitive data beyond the company network. As avowed by Tanjim, Oishi, Nandy, Jannah and Ahmed (2019), DLPs are critical cybersecurity components because they allow network administrators to determine the data types that could be transferred by end-users, as well as how they are transferred. The security incident and event management (SIEM) component has also been documented. In particular, SIEM is responsible for security event identification, monitoring, recording, and analysis in the network of an organization (Zhang, Qian, Wu, Hossain, Ghoneim and Chen, 2019). As such, their importance lies in the ability to inform network administrators regarding any anomalies, attacks, or breaches experienced in a corporation’s network. Lastly, there is a secure web gateway (SWG). Existing as a software solution or hardware appliance likened to a Firewall, SWG guards websurfing PCs against infection, ensuring further that company policies guiding the types of website users permitted to access the systems are enforced, as well as the types of downloads in which the users could engage (Liu, Shen, Cheng, Cai, Li, Zhou and Niu, 2018).

In this study, the technology on the focus involves cybersecurity, while the industry on the focus entails the software industry. The main aim is to discern the stages at which most software industries are in relation to the adoption and implementation of cybersecurity systems. Also, this secondary study aims to establish some of the beneficial effects that the software industry players have witnessed after implementing cybersecurity as a technology, as well as some of the challenges that have accrued from the same trend. The motivation is to predict the future implications of the perceived intersection. The analysis of data entails a content analysis approach.

In the findings, the software industry lags slightly when cybersecurity leaders are assessed. However, in the category of cybersecurity beginners, the software industry’s percentage is seen to be higher. From these affirmations, a resultant inference is that when the subject of the software industry’s usage, exploitation, or implementation of cybersecurity is investigated, most of the software companies are found to be at the beginning stage, with fewer firms emerging as leaders in cybersecurity technology implementation (or usage). The figure 2 below summarizes these observations.

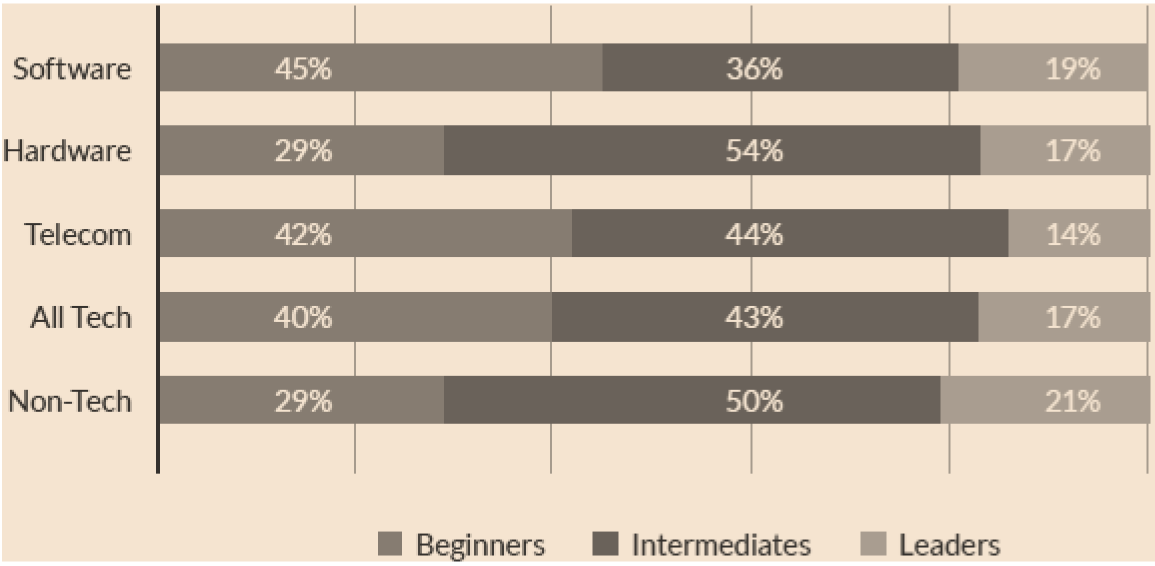

Figure 2: Level of Users and Usage

While software companies are expected to have exhibited heightened awareness regarding cyber threat evolution and led in cybersecurity usage, it is worth noting that resource and budgeting pressures associated with the cybersecurity function explain the trend in which most of the industry's players are at the beginner stage, rather than intermediate or leaders (Al-shukri, Lavanya, Sumesh and Krishnan, 2019). In particular, software companies are seen not only to focus on cybersecurity usage but also to compete with other initiatives and functions such as user experience improvement, digital transformation, and research and development (Farris, Taleb, Khettab and Song, 2019). Indeed, these multiple functions confirm the resource and budgeting pressures explain why many software companies, as reported in most of the recent scholarly studies, are at the beginner stage when the aspect of cybersecurity usage is evaluated (Bin Nordin et al., 2019). Giannakas, Papasalouros, Kambourakis and Gritzalis (2019) confirmed this observation by stating that there is a direct correlation between adequate cybersecurity funding and cybersecurity usage and function maturity in the software industry and that resource pressures imply that most of the firms are yet to allocate adequate funds to the cybersecurity functions, making most firms to remain (and operate) at the beginner stage – when it comes to cybersecurity usage.

The findings above demonstrate that an early stage of development characterizes most of the software companies' usage of cybersecurity. Additional scholarly studies have examined the subject of cybersecurity key functions at different stages of firm implementation and reported variations in the technology's perception based on the stage of implementation. For example, Haus, Orsag, Nunez, Bogdan and Lofaro (2019) focused on the perceived differences in software company executives' perception of cybersecurity usage depending on the cybersecurity maturity level and reported that in situations where firms are at the early-stage cybersecurity function, they are likely to emphasize specific functions such as reduced risks and incident prevention. On the other hand, the study demonstrated that in situations where software firms are at the leader stage, marked by advanced-stage cybersecurity functions, most of the executives tend to express a strategic function of cybersecurity, with most of the emphasis put on issues such as market share ability, customer engagement, and drivers of speed to market. As such, the findings suggest that the level of cybersecurity function and usage has not only seen many software companies remain at the beginner stage, but the level of cybersecurity function maturity accounts for differences in the functions that executives in the affected companies perceive of the technology, with most of the beginners emphasizing risk reduction and incident protection while most leaders emphasize market share, customer engagement, and cybersecurity functions as drivers of speed to market.

4. Conclusion

In summary, there are many the clustering algorithms such as KMeans, CMeans, Watersed and so on. KMeans has been shown to be very performant for image remote sensing clustering. In this study, we proposed two algorithms K_WIC and K_CIO. The algorithm K_WIC generates the good initialization center set and supports for speeding up execution of remote sensing image clustering. The algorithm K_CIO is used to cluster remote sensing images with adding context information of each pixel. The test results show that the algorithms K_WIC, K_CIO and K_WIC_CIO (a combination of K_WIC and K_CIO) can be used effectively for remote sensing image clustering. Besides, due to the nature of the Wavelet transform, the value domain of the output data is changed. Specifically, image doesn't belonging to the domain [0,255]. Therefore, we proposed an improvement of Wavelet transformation to still ensure that the domain of the output data belongs to the domain [0,255], suitable for image data. In future work, we will continue to study the new context information and the new algorithms.

5. REFERENCES

[1] Mallat S.G., *A theory for multi resolution signal decomposition, the wavelet representation*. IEEE transactions on Pattern Analysis and machine Intelligence, 11(7): 674-693, 1989.

[2] Muhammad Amir Shafiq and SaqibEjaz, *Real T ime Implementation of Multilevel Perfect Signal Reconstruction Filter Bank,* International Journal of Engineering & Technology IJET -IJENS Vol:10 No:04.

[3] NguyễnKhắcThờivàcộngsự, *GiáotrìnhViễnthám*, trườngĐạihọcNôngnghiệpHàNội, 2011.

[4] ShrutiDalmiyavàcộngsự, *Application of Wavelet based KMeans Algorithm in Mammogram Segmentatio*, International Journal of Computer Application, Volume 52– No.15, August 2012.

[5] Valliammal N., S.N.Geethalakshmi, *Leaf Image Segmentation Based On the Combination of Wavelet Transform and K Means Clustering*, International Journal of Advanced Research in Artificial Intelligence, Vol. 1, No. 3, 2012.

[6] Meritxell Bach Cuadra, Jean-Philippe Thiran, *Satellite Image Segmentation and Classification*, Diploma project, Fall 2004.

[7] Intanaidhayusoff, Nor ashidi mat isa, *Two-Dimensional Clustering Algorithms for Image Segmentation,* WSEAS Transactions on Computers, Issue 10, Volume 10, October 2011.

[8] Al-shukri, D., Lavanya, V., Sumesh, P. and Krishnan, P. (2019). Intelligent border security intrusion detection using IoT and embedded systems. *4th IEEE MEC International Conference on Big Data and Smart City (ICBDSC),* Muscat, Oman

[9] Farris, I., Taleb, T., Khettab, Y. and Song, J. (2019). A survey on emerging SDN and NFV security mechanisms for IoT systems. IEEE Communications Surveys & Tutorials, 21(1), 812-837

[10] Giannakas, F., Papasalouros, A., Kambourakis, G. & Gritzalis, S. (2019). A comprehensive cybersecurity learning platform for elementary education. *Information Security Journal: A Global Perspective*, 28:3, 81-106
[11] Hasan, R., Hussain, A., Nizamuddin, A. and Mahmood, A. (2018). An autonomous robot for intelligent security systems. *9th IEEE Control and System Graduate Research Colloquium (ICSGRC),* Shah Alam, Malaysia
[12] Haus, T., Orsag, M., Nunez, P., Bogdan, S. and Lofaro, D. (2019). Centroid vectoring for attitude control of floating base robots: from maritime to aerial applications'. *IEEE Access,* 7, 16021-16031
[13] Laaki, H., Miche, Y. and Tammi, K. (2019). Prototyping a digital twin for real time remote control over mobile networks: application of remote surgery. *IEEE Access,*7, 20325-20336
[14] Liu, K., Shen, W., Cheng, Y., Cai, L., Li, Q., Zhou, Q. and Niu, Z. (2018). Security analysis of mobile device-to-device network applications. *IEEE Internet of Things Journal, Early Access,* 1
[15] Tanjim, M., Oishi, A., Nandy, A., Jannah, R. and Ahmed, S. (2019). A flight control system for a vehicle. *IEEE International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST),* Dhaka, Bangladesh
[16] Uddin, S., Hossain, R., Rabbi, S., Hasan, A. and Zishan, R. (2019). Unmanned aerial vehicle for cleaning the high rise buildings'. *IEEE International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST),* Dhaka, Bangladesh
[17] Zanero, S. (2018). When cyber got real: challenges in securing cyberphysical systems. *IEEE Sensors*, New Delhi, India
[18] Zhang, U., Qian, Y., Wu, D., Hossain, S., Ghoneim, A. and Chen, M. (2019). Emotionaware multimedia systems security. IEEE Transactions on Multimedia, 21(3), 617-624
[19] Bin Nordin, M.N., Bin Mustafa, M.Z., Bin Abdul Razzaq, A.R. (2019). The practice of headmasters' leadership and its effect on job satisfaction of special education integration Program (PPKI) teachers in Johor, Malaysia. *Universal Journal of Educational Research*, 2019, 7(9), pp. 2008–2014