

Learning Activities Diagnostic Model Based on Educational Data Mining of Online Social Media Behavior

Khwunta KIRIMASTHONG* & Pakphoom PROMMOOL

Center of Excellence in AI and Emerging Technologies, School of Information Technology, Mae Fah Luang University, Thailand

*khwunta.kir@mfu.ac.th

Abstract: Due to the advancement of technology, the learning system has been rapidly changed from traditional learning through the materials of text and image only in the classroom-based learning to online-based learning using E-learning platform or MOOC (Massive Open Online Course). However, the existing online learning still have the limitation. For example, all learners learn from the same learning materials. Especially, the current online learning does not meet the needs and learning styles of learners in the 21st century that why the learners lack of motivation to study and fail to learn. Moreover, the social network such as Facebook, Twitter, YouTube have an influence to all people's everyday life. There are a big data of social media, learning media, human behaviors and interactions are generated rapidly in everyday. In the educational data analysis, the educators try to differentiate the learners based on their basic knowledge, needs and behaviors from social network data for offering them the appropriate learning activities. Therefore, the objective of this research is to propose the learning activities diagnostics model based on the educational data mining (EDM) that apply several data classification algorithms to analyze the learning behaviors. The learners' behaviors, reactions and interactions among them are collected from the simulated social network (Facebook) and the online learning ecosystem of the course of Introduction to Information Technology and Data Science, Mae Fah Luang University. This proposed learning activities diagnostic (LAD) model would lead us to indeed understand the nature of the learners and be able to provide the proper learning activities for both of the learners and instructors.

Keywords: Learning activities diagnostic, education data mining, online social media behavior, educational data classification

1. Introduction

Due to the advancement of technology, the learning system has been rapidly changed from traditional learning that is the classroom-based learning to the online-based learning. There are several online learning platforms such as MOOC, ThaiMooC, youtube, edX, Coursera, Udacity. Rai, L., and Chunrao, D. (2016) report that the success and failures in online learning is mostly depended on personal factors rather than factors influenced by the surroundings or the external environment. Moreover, Zhenghao et al. (2017) analyze the learning in Coursera, and reports that there is only 4% of all enrolled learners who has completed at least one online course. This research shows that the number of learners who has completed the online courses is less than the number of enrolled learners. The reason why most of learners could not complete the online courses that is because all learners have to study on the same online materials which may not relate to their interest. Moreover, they may get boring learning on un-interested learning materials and stop learning that course immediately, or they change to study other courses. The analysis of the influencing factors of success and failure in MOOC and general analysis of learner behavior (Rai, and Chunrao, 2016) found that the existing online learning platforms still have the limitation that are learners required more time for self-learning, un-interested or un-motivated learning materials that could not motivate the learner's self-learning accomplishment. Therefore, this research work will propose the learning activities diagnostics (LAD) model based on the educational data mining that apply several data classification algorithms to analyze the social media behaviors,

reactions and interactions among them from the virtual Facebook and the online learning ecosystem. This proposed LAD model would lead us to indeed understand the nature of the learners and be able to provide the proper learning activities for both of the learners and instructors.

2. Methodology

The proposed Learning Activities Diagnostic (LAD) model is developed based on the artificial intelligence to analyze the learning activities from our learning ecosystem and the personal behaviors from our virtual Facebook in Figure 1. The LAD model will be employed in the learning ecosystem to classify the learners' behaviors. Finally, the LAD model will classify personal online learning styles and then provides the learning activities that most suitable for each learner.

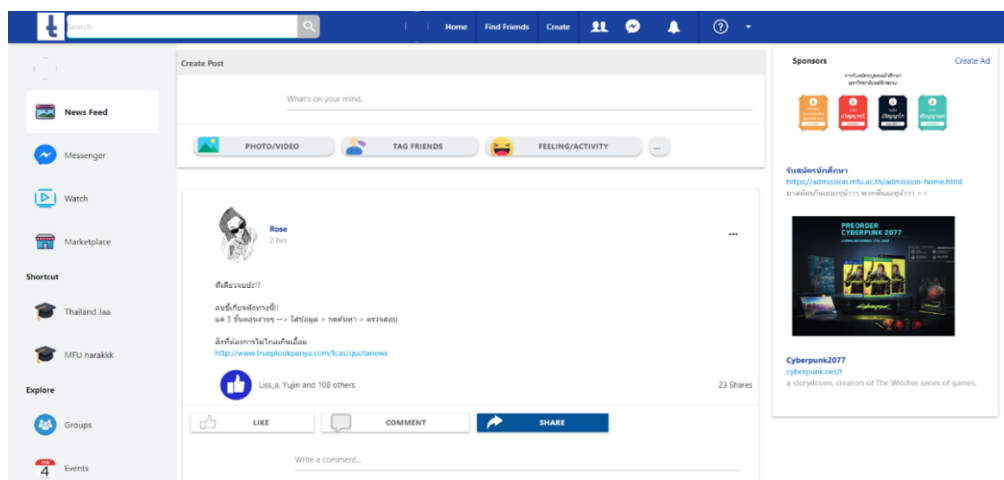


Figure 1. The virtual Facebook.

2.1 Experimental Dataset for Learning Activities Diagnostic Model

The LAD model is employed in the learning ecosystem that uses the application programming interface (API) for linking and extraction the learners' social media behaviors from the virtual Facebook and the learners' learning activities from that learning ecosystem. The experimental data of this LAD model including of two datasets that are the personal social behaviors of virtual Facebook and learning activities of learning ecosystem. The structure of the experimental dataset are shown in Table 1.

Table 1. The Structure of Training Dataset of LAD Model

student	From Facebook Data					From Online Learning Behavior												From LSI
	Personal Data	Personal Post				Module 1												Q
	gender (M / F)	1	2	3	4	5	Unit A (VIS/VER)	Unit A Time (s.)	Unit A practice (T/F)	Sequence (SEQ/GLO)	Unit B , C frist	Unit B Time (s.)	Unit B practice (T/F)	Unit C Time (s.)	Unit C practice (T/F)	Module 1 (ACT/REF)	Module 1 Intention (Y/N)	
0 (ศัลย์ช่าง)	M	video	video	img	img	status	VIS	40	T	GLO	C	75	T	90	F	REF	Y	1

The training dataset of LAD model in Table 1 for the learning style classification are from three main features. The first features from the virtual Facebook that are the personal data (gender) and the type of the last-five personal posts (link, offer, photo, status, or, video). The second features from the online learning behaviors thru the learning ecosystem. The third feature is the class of the eight-cluster online learning style index that are from the self-evaluation from the questionnaire based on the online learning style index in Figure 2 (Felder & Silverman, 1996; Hwang, et al, 2013; Pashler, et al, 2018) .

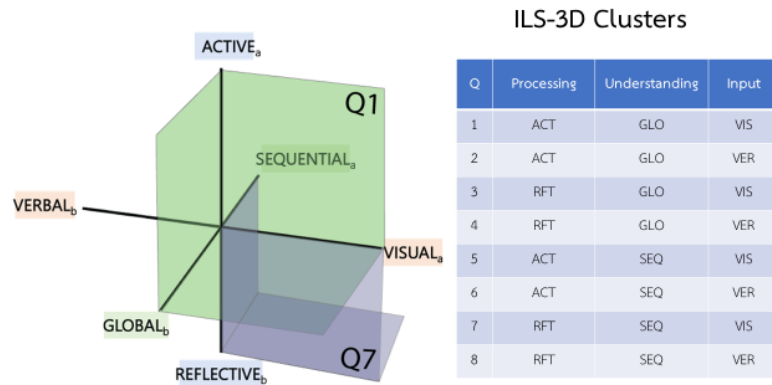


Figure 2. The eight-clusters of Online Learning Style Index.

2.2 Learning Activities Diagnostic Model

The learning activities diagnostic model is proposed to classify the learners' style based on their social media behaviors and their online learning scenario. This LAD model applies several classification techniques that are K-Nearest Neighbors, Decision Tree, Gaussian Naive Bayes, Support Vector Machine, and Multilayer Perceptron to analyze the training dataset combining with the learning style index (LSI) self-identification. Finally, the LAD model combined with the best classifier from this experiment will be used as the learning activities diagnostics to predict the proper class of learning style for new learners and then provide the learning activities that most matching with each learner. The proposed LAD model is illustrated in Figure 3.

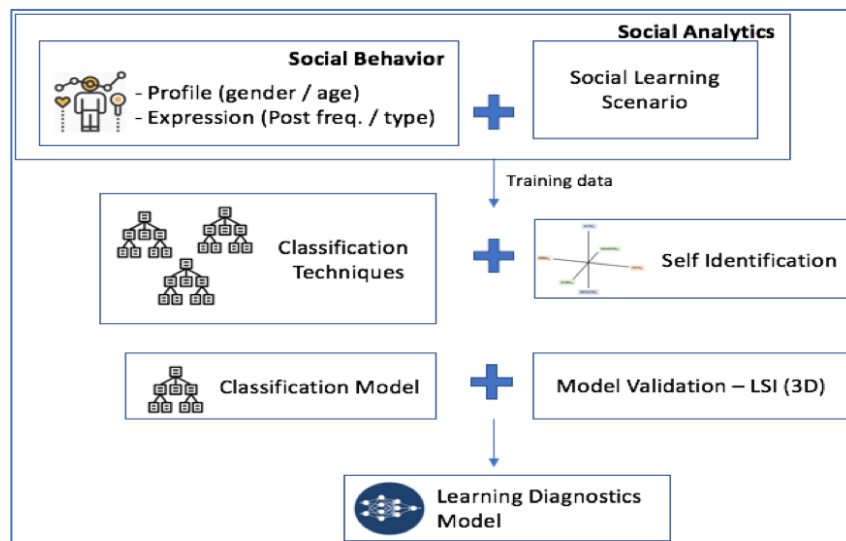


Figure 3. Learning Activities Diagnostic (LAD) Model.

2.3 Learning Activities Diagnostic Model Evaluation

This research work evaluates the performance of the proposed LAD classification by the cross validation. The confusion matrix of the classification result of LAD model in Table 2 is evaluated the cluster number of the Online Learning Style Index compared with the actual index of learning style from the questionnaire-based self-evaluation.

Table 2. *The Confusion Matrix*

LSI self-evaluation LAD classification	Q_{yes}	Q_{no}
Q_{yes}	True Positive	False Positive
Q_{no}	False Negative	True Negative

After that, the performance of the LAD model applied with different classification algorithm is evaluated in term of the accuracy, precision, recall, sensitivity and specificity shown as Equation (1) – (4).

$$\text{Sensitivity} = \frac{\text{True Positive}}{\text{All Disease (True Positive+ False Negative)}} \quad (1)$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive + False Positive}} \quad (2)$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive + False Negative}} \quad (3)$$

$$\text{Accuracy} = \frac{\text{True Positive+True Negative}}{\text{Total Data}} \quad (4)$$

3. Experimental and Results

This work tests the proposed LAD model applied with different five classification algorithm to the training dataset of a sampling group of 2,569 students. This sampling group is the first-year students from several faculties in Mae Fah Lung University who enrolled the course of Introduction to Information Technology and Data Science. Table 3 shows the performance of each proposed LAD model applied with different five classification algorithm. From the experimental results, we found that the LAD model combining with Decision Tree is the best classifier that gives the best accuracy 98.8%. Therefore, this work select the proposed LAD model that applies the Decision Tree as the classification algorithm to employ in our online learning ecosystem for further prediction the learning behaviors of new learners and to provide the proper learning activities for each learner.

Table 3. *The Evaluation of The LAD Model Applied with Different Classification Algorithms*

LAD model with classification algorithm	Performance (%)				
	<i>Sensitivity</i>	<i>Specificity</i>	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>
K-Nearest Neighbors	0.93	0.88	0.93	0.93	0.926
Decision Tree	0.99	0.98	0.99	0.99	0.988
Gaussian Naïve Bayes	0.98	0.98	0.98	0.98	0.985
Support Vector Machine	0.99	0.98	0.99	0.99	0.986
Multilayer Perceptron	0.98	0.98	0.98	0.98	0.987

After employing the proposed LAD model in the online learning ecosystem. The learning materials in our ecosystem is divided into four learning-modules. To test the LAD model employed in the learning ecosystem, all learners have to learn the same materials of module 1, and then the LAD model will analysis their social media behaviors from virtual Facebook and their online learning from the module 1 learning behaviors. From this experiment, the classification result of the LAD model combining with Decision Tree in Table 4 show that the learners' behaviors and interests are difference and are distributed in all eight clusters of the online learning style index. After the LAD model classifies the class of learning style index, the learning ecosystem will provide the materials of module 2-4 that exactly match with the predicted class of each learner. Therefore, our proposed LAD model employed with

the learning ecosystem will analyze and provide the suitable learning materials and activities for individual learner interests and learning style.

Table 4. *The Classification Result of The LAD Model Combining with Decision Tree*

(Q)	Online Learning Style						Nimer of Students			
	ACT	RFT	GLO	SEQ	VIS	VER	Male	Female	Total	%
1	✓		✓		✓		73	295	368	15.39
2	✓		✓			✓	63	163	226	9.45
3		✓	✓		✓		113	279	392	16.39
4		✓	✓			✓	66	135	201	8.41
5	✓			✓	✓		103	414	517	21.62
6	✓			✓		✓	67	296	363	15.18
7		✓		✓	✓		44	154	198	8.28
8		✓		✓		✓	30	96	126	5.27

4. Conclusion

From the experimental result shows that the proposed LAD model based on the analysis of the online social behaviors from virtual Facebook and the online learning behaviors from the learning ecosystem works well to classify the learning style of each learners. The proposed LAD model employed with the learning ecosystem will analyze and provide the suitable learning materials and activities for individual learner interests and learning style. Moreover, for the learners, the proposed LAD model will help them to gain more intention and interests in the online learning. For the instructors, this proposed model will help them to get more understanding in the learners' online learning behaviors and able to provide the learning activities and materials that matches with the learners.

References

- Felder, R. M., & Soloman, B. A. (1996). Index of learning style questionnaire. North Carolina State University.
- Rai, L., & Chunrao, D. (2016). Influencing factors of success and failure in MOOC and general analysis of learner behavior. *International Journal of Information and Education Technology*, 6(4), 262.
- Hwang, G., Sung, H., Hung, C., & Huang, I. (2013). A learning style perspective to investigate the necessity of developing adaptive learning systems. *Educational Technology and Society*, 16(2), 188–197.
- Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning Styles, Concept and evidence. *Psychological Science in the Public Interest*, 9(3), 105–119.
- Rai, L., & Chunrao, D. (2016). Influencing factors of success and failure in MOOC and general analysis of learner behavior. *International Journal of Information and Education Technology*, 6(4), 262.
- Zhenghao, C., Alcorn, B., Christensen, G., Eriksson, N., Koller, D., & Emanuel, E. J. (2017, November 30). Who's Benefiting from MOOCs, and Why. Retrieved from <https://hbr.org/2015/09/whos-benefiting-from-moocs-and-why>