Speily, O. R. B., & Kardan, A. A. (2018). Increasing Information Reposting Behavior in Online Learning Community. *Educational Technology & Society, 21* (4), 100–110.

Increasing Information Reposting Behavior in Online Learning Community

Omid R. B. Speily and Ahmad A. Kardan*
Department of Computer Engineering and Information Technology, Amirkabir University of Technology, Tehran, Iran // Speily@aut.ac.ir // aakardan@aut.ac.ir
*Corresponding author

(Submitted August 31, 2016; Revised January 5, 2017; Accepted March 18, 2017)

ABSTRACT

Online learning communities (OLCs) enable their learner to access different types of information through internet based structures anywhere anytime. OLCs are among the strategies used for the production and repost of information by learners interested in a specific area to support asynchronous learning. In this respect, learners become members of a particular domain and begin posting. OLC members consist of different sites with different educational backgrounds as well as different levels of expertise. This causes the sharing of posts which may not be appropriate for different learners. It also reduces data reposting behavior and subsequently decreases participation in information sharing. Furthermore, most learners of these communities take up a lurking position toward the posts. One of the ways proposed to increase information reposts is the selection and display of effective posts for each individual. Effective posts are selected in such a way that they can be more likely to be reposted by learners based on each learner's interests, knowledge and characteristics. The present paper intends to introduce a new method for selecting k effective posts to ensure the increase of information repost and participation in OLCs. In terms of participation in OLCs, learners are divided into two groups of posters and lurkers. Some solutions are proposed to encourage lurking learners to participate in content repostings. Comprehensive evaluations indicated that the proposed method had significantly solved the presented challenges.

Keywords

Asynchronous learning, Information reposting, Lurker, Online learning community

Introduction

Conventional teaching cannot meet the increasing need of people to learn due to inadequate resources and the limitations of time and space. Therefore, individuals should find and use other methods to learn more efficiently. In addition, learning development happens in places where the learners spend most of their time (Topping, 2005). Because of their reception by internet users, online communities and social media can be used for learning (Wagner, 2011). Currently, learners not only depend on conventional learning, but also use other learning environments on internet such as online learning communities (OLCs) to increase their learning opportunities (Ke & Hoadley, 2009). OLC is a kind of highly accessible and convenient learning platform where learners can search what interests them and share knowledge beyond the restriction of time and space. OLCs are computer-supported public or private groups (social networks) on the internet that address the learning needs of their members by facilitating asynchronous learning. Learners with special expertise help other learners needing knowledge or information. Despite their diversity, all these communities follow the same process: learner posts a content and other learners repost it if they likes it. These posts are virally shared on the network. Content sharing is an essential part of OLC experience. In addition to composing posts by themselves, learners can also rebroadcast or repost other learners' posts that they find of particular informational value. When a learner shares another learner's post, in fact, he is participating in the development of a common knowledge in his own network range. It is not only a feature to display his favorite posts. Repost of interesting posts has an extensive effect on networks and spreads information by exposing a new audience to the content (Shi, Rui, & Whinston, 2014).

First-stage use is an important indicator of OLCs' success, but the long-range success of OLCs lies on users' persistent-usage. Persistent-usage of OLCs can provide sufficient online learning materials and form prosperous online communication atmosphere which contribute to the long-term development of OLCs. The two fundamental interrelated challenges are access to appropriate information and shortage of participation (because of lurking behavior or low participation of some learner) in these communities. If information provided by a certain OLC can match users' information requirement, users are satisfied easily. User satisfaction plays a significant role in his/her participation and engagement (i.e., collaboration and sharing) which is vital for the OLCs' success. To address these challenges, this paper proposes a method to select the effective (appropriate) posts for each learner which increases the probability of repost behavior in the OLC.

Research questions addressed in this study are as follows:

ISSN 1436-4522 (online) and 1176-3647 (print). This article of the Journal of Educational Technology & Society is available under Creative Commons CC-BY-ND-NC 3.0 license (https://creativecommons.org/licenses/by-nc-nd/3.0/). For further queries, please contact Journal Editors at ets.editors@gmail.com

RQ1. How to predict effectiveness of the candidate post for each learner?
RQ2. Does the display of effective posts for various learners (lurker and poster) have an effect in increasing the reposting behavior and consequently participation of learners?

In the next section of this article, the literature on this subject are reviewed. Then, the proposed method is presented according to the nature of lurkers. The next section contains a detailed evaluation of the proposed method and the last section is devoted to summing up the study.

Literature review

Nowadays, online communities are extensively used and have induced fundamental changes to web-based systems. Therefore, according to the review of the performed studies, this paper is devoted to the role of online communities in learning as well as reposting behavior in online communities.

Over the last decade, with the development of web-based new application software on the basis of Web 2.0, social networks have received considerable attention. Social network services have provided this opportunity for the users to create online communities. Online communities improve interaction, information exchange, and personal experiences between users. The statistics reported in Duggan, Ellison, Lampe, Lenhart, and Madden's (2015) study show that in recent years the use of online communities has dramatically been increased in various parts of the world. Web 2.0 environment and tools such as messaging, e-mail, forum, wiki, social networks, and web conference are used in developing OLCs (Barczyk & Duncan, 2013; Lambić, 2016). OLCs are used outside the classroom for supporting the learning process. In addition, the use of social media in learning purposes is about to expand. Social media capabilities such as the ability in making synchronous or asynchronous connections, tagging, posting, creating and organizing virtual groups, and resource management and sharing have provided the possibility of easily implementation of OLCS (Mazman & Usluel, 2010). Learning with social media takes place with greater ease due to its reception capability and mobile nature in comparison with other platforms. Nowadays, many learners use social media for sharing information and knowledge, collaborating in conducting team projects, and discussing ideas and concepts (Dabbagh & Kitsantas, 2012). For example, it has been shown that learners used Facebook as a learning management system and it also satisfied them (Wang, Woo, Quek, Yang, & Liu, 2011) or based on (Chu & Meulemans, 2008) using social media are common between learners and 90% of information exchange between the learners happens via Facebook and MySpace.

In recent years, many issues regarding information sharing in OLCS have been proposed. According to (Deng & Tavares, 2013), information exchange has an important role in inspiring membership and activity in online communities and it also has a direct relationship with the value of an online community in the eyes of its members. In OLCS, sharing information improves the knowledge and skill of all group members. In such communities, sharing information increases the learner's tendency to participate and engage in the learning process (Junco, 2012). Information sharing is defined as an activity through which members exchange information, experience, and skill among themselves. Online communities are suitable for supporting interaction and sharing between the learners. Many studies have measured the impact of a special social media on information sharing and reposting behavior (Kleinberg & Ligett, 2013; Osatuyi, 2013). Other studies have mostly investigated effective factors in online community information reposting. Interaction and information sharing between members is considered to be the most important activity in OLC. Therefore, studying information reposting behavior and providing a method to increase its amount are important factors in OLCs' information sharing. Not many studies have been conducted on OLCs' information sharing using information reposting behavior. The study on blogsphere addresses "epidemic" interests among different blogs regarding the content cited or copied from other blogs (Adar & Adamic, 2005). By studying the cases, it estimates the relationship between two similar blogs. By relationship, it means the use of another post in the form of citation or copy. Another important point addressed in the study is the influence of a blog on another blog via a post. (Leskovec, Adamic, & Huberman, 2007) studied sharing small pieces of text (for example news) used in other articles and texts. For this purpose, a method was implemented by which the source of each piece could be specified in the network. This made it possible to study the structure of sharing in the network. The study aims to find the sources from which a post or posts are influenced. Given the existing studies and the nature of OLCs, the present article tries to present a method for displaying the best online information posts needed by learners leading to their increased participation. In this method, the importance of the displayed posts is taken into account in terms of post topic, learner interest, and information level in addition to the type of learner (poster or lurker).

Selection of *k* effective posts

The learner's profile is used to present a heuristic method for selecting effective posts for each learner. The proposed method needs no basic information on the learners because of easy-to-apply membership and using online communities (Pearson, Pearson, & Green, 2007). Only through the post content propagated in online communities does this method consider selecting effective posts for each learner. The learner's profile include the learner's connections, posts created or reposts from each learner. Due to the lack of a standard and diversity of implementation and design, in the proposed method the focus was only on the content assessment which is common in all kind of OLCs.

In order to choose the effective posts for each learner, this paper investigates three features: (1) subject and topic similarities of the selected post with the learner's interests, (2) The level of expertise of the author of the candidate post, (3) The novelty of the information based on the learner's information background. According to the learner's profile, favorite topics of the learner can be identified by the posts they create. Posts that are selected in line with the learner's interests are more suitable for them. Apart from interests, the learner's knowledge level differs in OLCS and each learner with different skill levels attempts to create a post. The skill level of each post differs according to the author's expertise. Although a post can be in line with the learner's interests, it becomes useless when it is far from the learner's level of expertise. Another factor which is taken into account in this paper is the novelty of information. Learning content, as a form of information posts, is useful for a learner when it is new to them. Moreover, the post's novelty differs from one learner to another. In the following section, a detailed explanation of the procedures is provided in order to measure these three factors. This section explains selection method of k effective posts in OLCs and discusses its characteristics.

Problem statement

Suppose that an OLC is implemented under a social network such as twitter. These groups are typically displayed as graphs of followers-followees. Learners follow other learners considering their interests and expertise. This directed graph is defined as $G = (V, E)$ where E represents the relationship between learners and V represents learners. Equation $(u, v) \in E$ shows that learner $\boldsymbol{u}$ follows learner $\boldsymbol{v}$. If $\boldsymbol{P}$ represents the total of posts created in the whole OLC, then an online social event "post" occurs when learner $\boldsymbol{u}$ creates post $\boldsymbol{p}$ at $t \in T$ time, represented as ***post(u, p, t)***. In the same manner, when learner $\boldsymbol{v}$ shares post $\boldsymbol{p}$ through learner $\boldsymbol{u}$ at time $\boldsymbol{t'}$, "reposting" occurs which can be represented as ***repost (v, u, p, t')***. According to the definitions provided, the probability of repost can be defined as a function of the probability of repost $\boldsymbol{p: P \times V \times V \times T} \rightarrow [1, 0]$. In this function, $\boldsymbol{T}$ is the temporal domain. The probable repost of $\boldsymbol{P}$ posts by any learner from $\boldsymbol{V}$ within the temporal domain $\boldsymbol{T}$ includes the values between zero and one (zero for not posting and 1 for posting).

If $\boldsymbol{\sigma}$ is taken as the selection procedure of $\boldsymbol{k}$ effective posts from candidate posts $(\boldsymbol{v, t})$ for learner $\boldsymbol{v}$ at any $\boldsymbol{t}$ time, the output $\boldsymbol{\sigma}(\boldsymbol{v, t})$ is the $\boldsymbol{k}$ effective post to learner $\boldsymbol{v}$ at time $\boldsymbol{t}$. The total candidate posts for learner $\boldsymbol{v}$ are those already created or reposted by learners set $\boldsymbol{V}$ followed by the learner $\boldsymbol{v}$. Equation (1) shows the initial set of candidate posts for $\boldsymbol{t' < t}$. From this stream of posts, duplicate posts already displayed for the learner in the previous time $\boldsymbol{t' < t}$ should be removed.

$$\text{Candid}(v, t) = \text{initial}(c, t) - \{p \in P | \text{post}(v, p, t') \cup \alpha(v, t')\} \quad (1)$$

Heuristic method for selection of *k* post

We have provided a heuristic method for the selection of *k* effective posts considering the computational capabilities and simplicity. This procedure is designed so as to be operational in online learning environments. In this section, the proposed method is introduced.

Similarity between posts (Post-Post)

To measure the similarities between posts, the method of text similarity is applied. For text similarity, the posts of a learner are considered as a set of word collection. According to the definition, this candidate post ($\boldsymbol{cp_i}$) is similar to the posts of learner $\boldsymbol{v}$ if it is related to his interests. $\boldsymbol{P_v = \{p_1, p_2, \ldots, p_n\}}$ is the collection of n posts created by learner $\boldsymbol{v}$. To determine the relationship between a candidate post $\boldsymbol{cp_i}$ and the topics of interest to learner $\boldsymbol{v}$, TF/IDF method and cosine of the angle between vectors of words $\boldsymbol{cp_i}$ and $\boldsymbol{P_v}$ have been used in many references. Owing to the diversity of the employed words, this method has low accuracy. For this purpose, different topics can be categorized in the texts using Latent Dirichlet Allocation (Bolelli, Ertekin, & Giles, 2009). Each topic includes a set of words $M_{topic\#} = \{w_1, w_2, \ldots, w_L\}$ for which the probability of the occurrence of L keywords is specified in the relevant topic. Assume two vectors of candidate posts cp_i and the previous posts of learner v ($P_v = \{p_1, p_2, \ldots, p_n\}$). Topic similarity is the angle between these two vectors which is measured through cosine using formula 2.

$$topsim(cp_i, P_v) = \frac{M_{topic_{cp_i}} . M_{topic_{P_v}}}{\|M_{topic_{cp_i}}\| . \|M_{topic_{P_v}}\|} \tag{2}$$

Novelty of a post

Novelty of a post is an important characteristic influencing the post's content value. On account of this definition, post p_i is novel to learner u_i if it is related to the interests of u_i while it is unknown to him. Methods have been introduced in (Liu, Ma, & Yu, 2001; Padmanabhan & Tuzhilin, 1999) for the calculation of unexpected information in retrieving web pages and Gaughan and (Gaughan & Smeato, 2005) and Gabrilovich, Dumais, and Horvitz, (2004) have proposed methods for the calculation of novel news. $P_{ui} = \{p_1, p_2, \ldots p_n\}$ is the set of posts written by learner u_i. As mentioned in the definition of novelty of a post, it is necessary to determine whether the post is related to the learner interests or not using LDA method (similar to previous section). Using LDA method, it is possible to determine the topic of each post in the form of a set of words ($M_{topic\#} = \{ w'_1, w'_2, \ldots, w'_L \}$) with the occurrence probability of each of them in the topic. Consequently, using this method, the topic of the post is identified and the class of words related to the topic is determined. If the stop words of candidate post cp_i are omitted using a standard method, and if $cp_i = \{w_1, w_2, \ldots, w_m\}$ is the words set used in post cp_i and $M_{topic\#} = \{ w'_1, w'_2, \ldots, w'_L \}$ is the words related to cp_i's topic, then equation (3) determines the post's novelty.

$$Novelty = \frac{topicsim(cp_i, P_{u_i})}{sim(M_{topic\#} - \{M_{topic\#} \cap cp_i\}, P_{u_i})} \tag{3}$$

As seen in equation (3), as topic similarity of cp_i and all u_i posts (numerator of formula 3) increases, the probability of novelty of the post increases. As mentioned in the definition of post novelty, it is necessary that the topic of cp_i is interesting for u_i. On the other hand, $M_{topic\#}$-{ $M_{topic\#} \cap cp_i$} is the set of words related to the topic of cp_i which are absent in the candidate post cp_i. In equation (3) sim($M_{topic\#}$-{ $M_{topic\#} \cap p_i$},P_{ui}) is the similarity of these words and all u_i posts. When this similarity is close to 1, it indicates that u_i has posted about the topic and related word of p_i and is probably familiar with this post. So the probability of novelty of the post p_i declines. As seen in equation (3), as the similarity of topic-related words to the learner's topics of discussion increases, the probability of the novelty of the post declines. Post novelty is directly related to its similarity to the learner's posts by definition.

Similarity between learners (Learner - Learner)

People with different expertise share their posts on the network. It is very beneficial to find learners with common fields. For both learners u, v ∈ V, the degree of similarity is equal to the degree of similarity between the posts already created. The essential thing about sharing information in OLCs is to find people with the same level of information in addition to similar posts. For example, a learner who has created more than 100 posts about smart phone applications is different from someone who has just had a few posts or reposts in the same field. For either learner, the action vector can be defined (Equation 4). This vector contains n keywords created or reposted by the learner v ∈ V. Weighted cosine is used to determine the similarity between the vectors of learners. In this respect, the coefficients i (the number of keyword repeated by learner v ∈ V) is determined for n keywords till time t. Considering coefficients i, the level of learners' knowledge on a specific area is determined according to the number of posts made by them. The two learners are examined and taken into account for determining the similarity given the repetition of the keywords in the posts.

$$vector_u^t = i_1^t keyword_1 + \ldots + i_n^t keyword_n \tag{4}$$

The degree of similarity between learners u and V is equal to the value of cosine for the vectors of these two learners.

$$sim(u, v) = \cos(vector_u, vector_v, t) = \frac{vector_u^t . vector_v^t}{\|vector_u^t\| . \|vector_v^t\|} \tag{5}$$

In equation (5), $\|vector_v^t\|$ or/and $\|vector_u^t\|$ represent the value of action vector for learners ***v*** and ***u,*** respectively**.** If any of these values is zero, it means that the relevant learner has had no action (neither created nor reposted). In that case, the similarity between two learners is not defined. Accordingly, in collecting data, only those learners are taken into account that have at least created 10 posts or reposted. It should be noted that the action vector of a learner changes with time. In this respect, the action vector of learners at time ***T*** is used in each determination of similarity between two learners.

Using logistic regression to predict the probability of reposting

Logistic regression is used to estimate the probability of post effectiveness for each learner. The binary logistic model is used to estimate the probability of a binary response (reposts happened or not) based on one or more features (similarity of post-post, post novelty, and similarity between learners).

The goal of logistic regression in this problem is to find the best fitting model to describe the relationship between the effectiveness of the post (dependent variable, response or outcome variable) and a set of independent variables (3 features) using the sigmoid (logistics) function (Equation 6).

$$y_i = h(w^T x_i) = \frac{1}{1+\exp(-w^T x_i)} \quad (6)$$

In this equation, (y_i = 1 for effective posts (repost happened) and y_i = 0 for otherwise) y_i is the prediction based on x_i inputs (value of each features). w^T is the vector of coefficients of each feature obtained from training data. In addition to classification, this method also has probable output. For example, $h_w(x_i)$ = 0.8, means that the probability of post effectiveness is 80% for the sample ($h_w(x_i) = p(y_i = 1 | x_i; w^T)$).

Probability of post effectiveness based on learner type (Lurker & Poster)

In the proposed method, in addition to the structural characteristics of learners in the network, the attention is paid to the attitude of reposting among learners. Learners are divided into two categories of posters and lurkers based on their reposting attitude (Sloep & Kester, 2009). Lurkers are those who become a member of an online educational community but do not post, are only readers, and are not active (Nonnecke & Preece, 2000; Nonnecke & Preece, 2001). Considering this type of learners, the present study intends to reduce pure lurking behavior by setting the time for online access to posts based on the importance of that post; that is, the posts which are more likely to be reposted by learners, are displayed longer for lurkers than for posters. The calculated probability of repost of post ***p*** of learner ***u*** by learner ***v*** at timestamp ***t*** (put formally ***p*(*repost* (*v, u, p, t*))** is shown in table (1). In algorithm (1), t_u is reposting time by learner ***u. t*** is the time of running the algorithm and γ_v is the average time interval between the posts of learner ***v.*** α is the adjusted coefficient of γ_v. At the zero line of this algorithm, by calculating the elapsed time since the repost of the posts by learner ***u***, if time is less than the average time interval between the posts of learner ***v*** (γ_v), there is still the probability that learner ***v*** has not seen post ***p.*** In this respect, the probability of reposting by learner ***v*** is equal to max($p_{u,v}^p, \varepsilon$).Line 2 evaluates a condition in which the time elapsed since the repost of the posts by learner ***u*** is longer than the average time interval between the posts of learner ***v*** (γ_v). In this case, the learner had most likely seen the post but was unwilling to repost it. Considering this fact, it is least probable that learner ***v*** reposts the post ***p*** of learner ***v*** in timestamp ***t*** (***p(repost* (*v, u, p, t*))**. ε is equal to minimal value at 10^{-3}.

Table 1.Algorithm:calculating the probable repost of post p of the learner ***u*** by the learner ***v*** at timestamp ***t***

00	If $t - t_u <= \alpha\gamma_v$
01	***p*(*repost* (*v, u, p, t*))**=max($p_{u,v}^p, \varepsilon$)
02	If otherwise
03	***p*(*repost* (*v, u, p, t*)) = ε**
04	End

The coefficient of α provides more opportunity for lurking learners by adjusting the impact of γ_v. Normally, γ_v of lurking learners is far longer than γ_v of other learners. For ease of calculation, the coefficient α of γ_v is considered to be 1 and 2 for poster and lurker learners, respectively. In this respect, when a learner is lurking, he has more time to read and repost the posts with higher probability.

Methods

Data collection

The data presented in this study were collected based on an unrestricted self-selected web survey (Fricker, 2008). Following the methodological procedure for posting surveys on discussion boards (Ip, Barnett, Tenerowicz, & Perry, 2010), a survey invitation was published in forum with the prior approval of the online community managers. The survey was piloted between March 12 and July 12, 2015. The participants were selected from information technology and electrical engineering technophiles in popular Iranian online communities (online communities are similar to twitter environment and its members can follow other members, interested topics, etc.). Among them, only those learners are taken into accounts that had at least created 10 English posts or reposts and had more than 10 items in their list of followers and followees. In this method, qualified members were selected for a 10-day assessment period. A total of 5 to 20 posts a day were prepared for the volunteers' assessment. Volunteers participated in the assessment independently on a daily basis without knowing other users' opinions. These volunteers examined the posts to repost them and answered specific questions regarding each proposed feature of the posts (novelty and relevance). They ranked questions on novelty of the posts and relevance of the posts to their interests in the range of 1-5. In the 1 to 5 range of answers, 1 shows the lowest score of the feature in question, while 5 shows the highest score of that feature. Among the participants, lurkers are those who do not (re)post any posts publically in a certain period of time (Nonnecke & Preece, 2000) or those who rarely participate in the study (Sloep & Kester, 2009)'s (2009). Like the method presented in (Amichai-Hamburger et al., 2016), learners were asked about the last time they had (re)posted a post to identify lurkers. Based on their last (re)post in online communities, learners can select from more than one day, one month, one year and never. In this study, learners who selected more than one year or never were considered to be lurkers.

Sample

Non-probability sampling method was used to select more than 790 participants. In total, 608 participants participated in the study voluntarily and free of charge. Then, 583 participants were found acceptable. Therefore, 263 participants (45%) were selected as posters from 583 respondents, and 320 participants (55%) were selected as lurkers. The average age of lurkers was lower than posters; however, there was not a significant difference. According to statistics, 63% of respondents were male, and 37% were female. The majority of respondents were in 20-30 (34%) and 30-40 (41%) year old groups. The membership duration in online communities of 94% of respondents was more than 6 months; a fact which indicates the familiarity of most of them with the environment of an online community. Among them, 63% had academic educations (21% with master's and PhD degrees) and 37% did not have academic educations. Almost 79% of them were married, and the rest were single.

Analysis

Evaluating methodology was consisted of three phases. In the first phase, for each of the participants, ***k*** posts (as the learner wished) were displayed on a daily basis. Learners decided about the post's reposting when they observed them and they also determined when to perform the test during the day. In the second phase, according to their repost decision (previous phase), ***k*** effective and proper posts were displayed and like the first phase, learners decided about the post's reposting. In this phase, the prediction of reposting behavior was assessed. Three well-known measures, namely precision, recall, and F measure, were used to assess the predictions (Tang, Miao, Quan, Tang, & Deng, 2015; Zhang et al., 2012). These measures are used for prediction problems of binary class (posting or not reposting). Equations 7 to 9 show these measures.

$$Precision = \frac{|\{\text{Predicted RT}\} \cap \{\text{True RT}\}|}{|\{\text{Predicted RT}\}|} \tag{7}$$

$$\text{Recall} = \frac{|\{\text{Predicted RT}\} \cap \{\text{True RT}\}|}{|\{\text{True RT}\}|} \tag{8}$$

$$F = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{9}$$

To evaluate the solutions for estimated relevance and novelty of a post, Kendall's rank correlation coefficient was used to compare the measured values of both features (post novelty and relevance to the user) with the user-expressed values (1-5 range). This method is employed to obtain the correlation between the ranks of two quantities. If y_i is the value expressed for each feature (novelty and relevance) by the user, and x_i is the proposed measured value by the proposed solution and (x_1 , y_1), (x_2, y_2), …, (x_n, y_n), then the values obtained for all of the n post are considered. Each (x_i , y_i) and (x_j , y_j) pair ($i \neq j$) is known as a concordant pair if $x_i > x_j$ then $y_i > y_j$ or if $x_i < x_j$ then $y_i < y_j$. The pair is discordant if $x_i > x_j$ then $y_i < y_j$ or if $x_i < x_j$ then $y_i > y_j$. If the two quantities are equal, the pair is neither concordant nor discordant. Kendall's correlation coefficient is calculated by equation (10), and varies between 1 and -1 ($-1 \leq \tau \leq 1$). If the rank of a post based on the aforementioned two features is close to its rank based on user opinions (in the questionnaire), this coefficient will be closer to 1. If the ranks differ completely, this coefficient will be closer to -1, whereas 0 shows independence of the two values.

$$\tau = \frac{(\text{number of concordant pairs}) - (\text{number of discordant pairs})}{\frac{1}{2}n(n-1)} \tag{10}$$

The correlation of these features with repost behavior is studied in this paper. If the values of these features are significantly related to repost behavior, the relationship can be measured based on conventional learning. To this end, a method similar to Pearson's correlation method is employed. Since each feature has continuous values and repost behavior is a binary variable (0 or 1), Pearson's method cannot be used. Point-biserial method is utilized for such problems. If the values of each feature of each post are a continuous variable (x), and repost behavior for the post is a binary variable (y), then the point-biserial correlation coefficient is based on equation (11), where M_1 shows mean feature value for posts, leading to repost behavior y=1. Similarly, M_0, is the mean of features of posts not leading to repost behavior $y = 0$. Moreover, n_1 shows the number of posts in the samples except for posts leading to repost behavior ($y = 1$), and n_0 denotes the number of posts in samples which do not result in repost behavior ($y = 0$). In addition, n is the total number of samples examined, s_n is the standard deviation for values of features of all studied post samples. This correlation coefficient varies between -1 and 1, where 1 shows maximum positive correlation between measured values and user behavior and -1 shows maximum negative correlation between measured values and user behavior. Zero (0) also shows independence of the features from user repost behavior.

$$r_{pb} = \frac{M_1 - M_0}{s_n} \sqrt{\frac{n_1 n_0}{n^2}} \tag{11}$$

In the last phase, to evaluate the increasing of information repost behavior in OLC, 5 more methods other than the proposed one, were also employed: (1) latest post method: based on the time a post was created and learners visited the latest ***k*** post, (2) random ***k*** post selection method for a learner, (3) post-post similarity method: ***k*** posts with the highest similarity to the learner's posts, (4) top ***k*** novel posts: ***k*** post selection method with the highest measured degree of novelty, (5) learner-learner similarity method: ***k*** post selection method with the highest similarity of the posts with the learner. Furthermore, learners were categorized into lurker and posters. Each of these categories is randomly divided into 6 categories. During the 10 day evaluation, each of these 6 categories was received ***k*** post alternatively without advanced notice based on one of these six methods. Equation (12) shows the increasing reposting behavior evaluation metric.

$$repost\ rate = \frac{Number\ of\ reposts}{Number\ of\ visited\ Posts} * 100 \tag{12}$$

Results

Table (1) shows the evaluation metrics of the proposed repost behavior prediction using logistic regression for lurkers and the posters. The results showed that the proposed method had precisely predicted the learner's reposting behavior. In order to assess the accuracy of choosing k effective post, common criteria were used for estimation evaluation.

Table 1. Repost behavior prediction evaluation

	Posters			Lurkers		
	F measure	Recall	Precision	F measure	Recall	Precision
Proposed Repost Prediction	0.643	0.541	0.793	0.554	0.477	0.662

Table (2) presents the results of analysis of Kendall's correlation coefficient for all posts in the set. Figures 1 and 2 indicate the scatter plot of the estimated novelty and relevance ([0,1]) and observed novelty and relevance (range 1 to 5).

Table 2. Kendall's tau correlation coefficient

	Relevance (Average of P-P & L-L similarity)	Novelty
τ	0.68	0.54

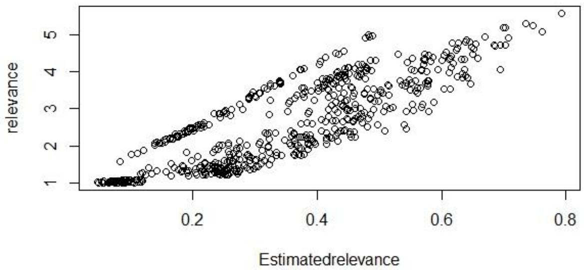

Figure 1. Estimated Relevance (range 0 to 1) Vs. Relevance (questionnaire 1-5)

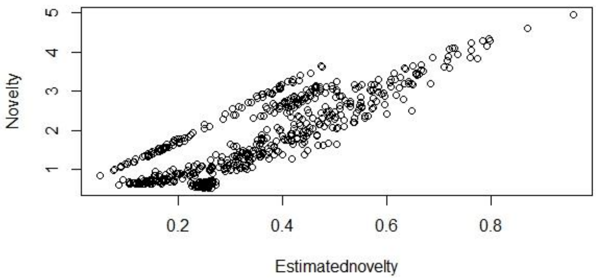

Figure 2. Estimated Novelty (range 0 to 1) Vs. Novelty (questionnaire 1-5)

Table 3 presents the results of correlation coefficient between features and repost behaviors. These results are reflective of a positive correlation between these features and repost behavior. The maximum correlation belonged to post-post similarity and topical similarity, whereas the minimum correlation belonged to novelty of posts.

Table 3. Coefficient of correlation between proposed features and repost behavior

	Post-Post Similarity	Novelty	Learner-Learner similarity
r_{pb}	0.63	0.49	0.54

According to Table 3, using the proposed method the average of repost rate for each of these two categories, namely, lurkers and posters, was more than those of 5 other methods.

Table 4. Repost rate for each of methods

Methods	Posters	Lurker
	Repost Rate	Repost Rate
Random	11	4
Recent	7	9
Post-Post Similarity	34	22
Learner-Learner Similarity	28	17
Post Novelty	24	20
Proposed Method	42	34

Also, in order to have a better comparison, a set of one-tailed *t*-tests were taken on 10-fold cross validation results. In this experiment, we considered the result of the proposed method as algorithm A2. Hence, if the result of t-test between the proposed method and another algorithm is a negative value, then the proposed method statistically outperforms the other algorithm. The statistical differences between the experiments obtained with the proposed method, and the other algorithm for the test data sets are listed in Table 5.

Table 5. The result of statistical *t*-test between the proposed method and other methods

	Poster	Lurker
Random	-3.924	-5.342
Recent	-79.402	-34.34
Post-Post Similarity	-29.931	-6.629
Learner-Learner Similarity	-6.456	-15.686
Post Novelty	-30.713	-36.554

Discussion

Generally, the topic of increasing the repost of information has been rarely studied in education-related studies despite its great importance in social networks. In this study, in order to answer research question 1, prediction of the effectiveness of the post from the content assessment of candidate posts was emphasized regarding the variety of OLCs and their various capabilities. Features related to the relevance (post-post and learner-learner similarity) and novelty of posts were considered for the prediction of post effectiveness (see Table 1). In research regarding the correlation of relevance and novelty features, a comprehensive assessment was done and the results of the direct effect of these features on learners' behaviors are shown in Table 3. The results of this study were in line with the results found in various previous research studies such as (Lee & Lee, 2009; Sarma & Panigrahy, 2010; Wang, Duan, Koul, & Sheth, 2014). In addition, the measurement methods of the presented features (relevance and novelty of the post) of the study were compared through the explicit feedbacks of learners in the form of a questionnaire and their accuracies were also assessed (see Table 2).

The role of the proposed method in increasing repost behavior and consequently online participation (the answer to research question 2) was assessed and confirmed for both lurker and poster learners.

Regarding research question 2, both categories of learners had positive repost rates through the selection of effective posts and the display method of this study. The increase in reposts leads to an increase in the participation of learners in online society and sharing of more information. The significance of the improvement through the suggested method compared with other methods is proven in Tables 4 and 5.

Limitation and future research opportunities

The most obvious limitation of this study was the lack of a specific standard in the implementation and design of OLCs. Therefore, the suggested method was designed based on the content of posts to make it applicable in various kinds of OLCs. Other factors such as external events (such as trends and news), context of OLC (such as health forums or technical forums), security concerns and OLC design problems can be effective in learners' repost behavior. The investigation of these factors is ignored due to the lack of a specific standard and high quality data. Considering at least two kinds of OLCs, future studies can investigate more effective factors in learners' repost behavior to increase the accuracy of the suggested method. In addition, the display of effective posts under adaptive navigation can be presented for learners based on an educational goal.

Conclusion

OLCs are public or private groups on the internet that address the learning needs of their members by facilitating asynchronous learning. This paper seeks to solve the two fundamental linked challenges being access to the appropriate information and shortage of participation (because of lurking behavior or low participation of some learner) in these communities by selecting ***k*** effective posts for learner ***v*** among all candidate posts. These posts should be selected based on learner interests, characteristics and knowledge level to motivate the learner to repost it. This reposting behavior facilitates information sharing and increases participation in online communities. For online implementation, the proposed method is suitable to be uses in all kinds of OLCs. Certain features such as post-post similarity, learner-learner similarity, and post novelty are among major parameters taken into account in the selection of ***k*** effective posts. The proposed method focuses on the problem of online communities that is lurking learners. Comperhenisve evaluations indicate that the proposed method significantly outperfomed the existing methods.

Reference

Adar, E., & Adamic, L. A. (2005). Tracking information epidemics in blogspace. In *Proceedings of the 2005 IEEE/WIC/ACM international conference on web intelligence* (pp. 207-214). Washington, DC: IEEE Computer Society.

Amichai-Hamburger, Y., Gazit, T., Bar-Ilan, J., Perez, O., Aharony, N., Bronstein, J., & Dyne, T. S. (2016). Psychological factors behind the lack of participation in online discussions. *Computers in Human Behavior, 55*, 268-277.

Barczyk, C. C., & Duncan, D. G. (2013). Facebook in higher education courses: An Analysis of students' attitudes, community of practice, and classroom community. *International Business and Management, 6*(1), 1–11.

Bolelli, L., Ertekin, Ş., & Giles, C. L. (2009, April). Topic and trend detection in text collections using latent dirichlet allocation. In *European Conference on Information Retrieval* (pp. 776-780). Berlin, Germany: Springer.

Chu, M., & Meulemans, Y. N. (2008). The Problems and potential of MySpace and Facebook usage in academic libraries. *Internet Reference Services Quarterly, 13*(1), 69-85.

Dabbagh, N., & Kitsantas, A. (2012). Personal learning environments, social media, and self-regulated learning: A Natural formula for connecting formal and informal learning. *The Internet and higher education, 15*(1), 3-8.

Deng, L., & Tavares, N. J. (2013). From Moodle to Facebook: Exploring students' motivation and experiences in online communities. *Computers & Education, 68*, 167-176.

Duggan, M., Ellison, N. B., Lampe, C., Lenhart, A., & Madden, M. (2015). Social media update 2014. *Pew research center, 9.*

Fricker, R. D. (2008). Sampling methods for web and e-mail surveys. In *The SAGE handbook of online research methods* (pp. 195-216). London, UK: Sage.

Gabrilovich, E., Dumais, S., & Horvitz, E. (2004, May). Newsjunkie: providing personalized newsfeeds via analysis of information novelty. In *Proceedings of the 13th International Conference on World Wide Web* (pp. 482-490). New York, NY: ACM.

Gaughan, G., & Smeaton, A. F. (2005, October). Finding new news: Novelty detection in broadcast news. In *Asia Information Retrieval Symposium* (pp. 583-588). Berlin, Germany: Springer.

Ip, E. J., Barnett, M. J., Tenerowicz, M. J., & Perry, P. J. (2010). The Touro 12-step: A Systematic guide to optimizing survey research with online discussion boards. *Journal of medical Internet research, 12*(2). doi:10.2196/jmir.1314

Junco, R. (2012). The Relationship between frequency of Facebook use, participation in Facebook activities, and student engagement. *Computers & Education, 58*(1), 162-171.

Ke, F., & Hoadley, C. (2009). Evaluating online learning communities. *Educational Technology Research and Development, 57*(4), 487-510. doi:10.1007/s11423-009-9120-2

Kleinberg, J., & Ligett, K. (2013). Information-sharing in social networks. *Games and Economic Behavior, 82*, 702-716.

Lambić, D. (2016). Correlation between Facebook use for educational purposes and academic performance of students. *Computers in Human Behavior, 61*, 313-320.

Shin, H., & Lee, J. (2009, April). Ranking user-created contents by search user's inclination in online communities. In *Proceedings of the 18th international conference on World wide web* (pp. 1215-1216). New York, NY: ACM.

Leskovec, J., Adamic, L. A., & Huberman, B. A. (2007). The Dynamics of viral marketing. *ACM Transactions on the Web (TWEB), 1*(1), Article No. 5. doi:10.1145/1232722.1232727

Liu, B., Ma, Y., & Yu, P. S. (2001, August). Discovering unexpected information from your competitors' web sites. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 144-153). New York, NY: ACM.

Mazman, S. G., & Usluel, Y. K. (2010). Modeling educational usage of Facebook. *Computers & Education, 55*(2), 444-453.

Nonnecke, B., & Preece, J. (2000, April). Lurker demographics: Counting the silent. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 73-80). New York, NY: ACM.

Nonnecke, B., & Preece, J. (2003). Silent participants: Getting to know lurkers better. In *From usenet to CoWebs* (pp. 110-132). London, UK: Springer.

Nonnecke, B., & Preece, J. (2001). Why lurkers lurk. *In Proceedings of Americas Conference on Information Systems (AMCIS)* (pp. 1521-1530). Retrieved from http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1733&context=amcis2001

Osatuyi, B. (2013). Information sharing on social media sites. *Computers in Human Behavior, 29*(6), 2622-2631.

Padmanabhan, B., & Tuzhilin, A. (1999). Unexpectedness as a measure of interestingness in knowledge discovery. *Decision Support Systems, 27*(3), 303-318.

Pearson, J. M., Pearson, A., & Green, D. (2007). Determining the importance of key criteria in web usability. *Management Research News, 30*(11), 816-828.

Das Sarma, A., Das Sarma, A., Gollapudi, S., & Panigrahy, R. (2010). Ranking mechanisms in twitter-like forums. In *Proceedings of the third ACM international conference on Web search and data mining* (pp. 21-30). New York, NY: ACM.

Shi, Z., Rui, H., & Whinston, A. B. (2013). Content sharing in a social broadcasting environment: evidence from twitter. *MIS Quarterly, 38*(1), 123-142.

Sloep, P., & Kester, L. (2009). From Lurker to active participant. In *Learning network services for professional development* (pp. 17-25). Berlin, Germany: Springer.

Tang, X., Miao, Q., Quan, Y., Tang, J., & Deng, K. (2015). Predicting individual retweet behavior by user similarity: A Multi-task learning approach. *Knowledge-Based Systems, 89*, 681-688.

Topping, K. J. (2005). Trends in peer learning. *Educational Psychology, 25*(6), 631–645.

Wagner, R. (2011). Social media tools for teaching and learning. *Athletic Training Education Journal, 6*(1), 51–52.

Wang, Q., Woo, H. L., Quek, C. L., Yang, Y., & Liu, M. (2011). Using the Facebook group as a learning management system: An Exploratory study. *British Journal of Educational Technology, 43*(3), 428-438.

Wang, W., Duan, L., Koul, A., & Sheth, A. P. (2014, June). *YouRank: Let user engagement rank microblog search results.* Paper presented at the Eighth International AAAI Conference on Weblogs and Social Media, Ann Arbor, MI.

Zhang, H., Zhao, Q., Liu, H., He, J., Du, X., & Chen, H. (2012, November). Predicting retweet behavior in Weibo social network. In *International Conference on Web Information Systems Engineering* (pp. 737-743). Springer, Berlin, Heidelberg.