MIS Review

Vol. 22, Nos. 1/2, 2017

GPN 2007800019
ISSN 1018-1393

MIS Review

An International Journal

Vol. 22, Nos. 1/2, 2017

E-ISSN:2218-3450 / ISSN:1018-1393 / GPN:2007800019

Published by Department of Management Information Systems,

College of Commerce, National Chengchi University

MIS Review **Vol. 22, Nos. 1/2** Published in September (2016)/March (2017)

Publisher: Eldon Y. Li

Published by: National Chengchi University, Department of Management Information Systems

Editor-in-Chief: Eldon Y. Li

Executive Editor: Rua-Huan Tsaih

Assistant Editor: Laurence Fang-Kai Chang

Editorial Assistant: Marian Y.S. Chan

Publication Office:

National Chengchi University, Department of Management Information Systems No. 64, Sec. 2, ZhiNan Rd., Wenshan District, Taipei City 116, Taiwan

E-ISSN: 2218-3450

ISSN: 1018-1393

GPN: 2007800019

Editorial Board

MIS Review

September (2016) /March (2017) Vol. 22, Nos. 1/2

Contents

Editor's Introduction

In this MISR issue, we are delighted to present four research papers. The summary of the four papers is as follows.

Te-Min Chang, Wen-Feng Hsiao, and Ming-Fu Hsu in their paper "Hidden Topic Analysis for Personalized Document Recommendation" propose hybrid filtering approaches for document recommendation system to overcome the shortcomings of collaborative filtering (CF) and content-based filtering (CBF) alone. Particularly, they incorporate latent Dirichlet allocation (LDA) to uncover latent semantic structure in the collected document corpus. The hidden topic results can act as the technical bridge between CBF and CF because we can either obtain robust document similarity in CF, or to further explore user profiles in CBF. Two experiments are conducted accordingly. The results show that their proposed approaches outperform other counterparts on the recommendation performance, which justifies the feasibility and practical applications of the proposed approaches.

Asad Ahmad and Mohammed Naved Khan in their paper "Factors Influencing Consumers' Attitudes toward Social Media Marketing" state that people with common interest interact and share their experiences with the other members on the Social Networking Sites (SNS). SNSs have become a centre stage for e-commerce with the phenomenal rise in the number of the SNSs users. It has become a platform where different marketers advertise their products and services. This medium has equipped the marketers in reaching their target consumers easily. The present study uses an 18 item scale to explore the factors determining attitude of the students towards advertisement over SNSs. Usefulness, Reliability and Word of Mouth Quality emerge to be the critical factors determining the attitude of the SNSs users towards the advertisements.

Abdul Qadir and Swati Agrawal in their paper "Human Resource Information System (HRIS): Re-engineering the Traditional Human Resource Management for Leveraging Strategic Human Resource Management" argue that today's Human Resource Management (HRM) has travelled from attendance sheet to balance sheet of the organization in the pursuit of transforming HRM to Strategic Human Resource Management (SHRM). In the fast growing business world, the changing paradigms have pushed HRM functions to revive and deliver on the business front. Human Resource Information System (HRIS) has leveraged SHRM in more than one ways for HR practitioners enabling the transformation of traditional HRM to transformational HRM. This paper presents a conceptual framework through an HRIS model underpinning the potentials and opportunities that an HRIS offers for practicing SHRM in the 21st century for the academia and industry professionals.

Othmar Mwambe and Isao Echizen in their paper "Security Modeling Tool for Information Systems: Security Oriented Malicious Activity Diagrams Meta Model Validation" argue that the dramatic growth of information system security attacks remains a nightmare. They use various information system security risk management approaches and security modelling languages to address information system security threats. Among many other security modelling languages, Mal-Activity Diagrams

(MAD) have been used to model system malicious processes and risk mitigation processes. However, due to their syntactic and semantic drawbacks, Security Oriented Malicious Activity Diagrams (SOMAD) were introduced in their previous study as an extension of MAD. In this study, they use industrial survey to validate the comprehensiveness and applicability of SOMAD Meta model. The results show that SOMAD Meta Model is a comprehensive tool enough to address information system security issues at large scope.

As the final note, we would like to thank all the authors and reviewers for their collaborative efforts to make this issue possible. It is our sincere wish that this journal become an attractive knowledge exchange platform among information systems researchers. Last but not least, to our loyal readers around the world, we hope you find the contents of the papers useful to your work or research.

Eldon Y. Li

Eldon Y. Li, Ph.D.
Editor-in-Chief and University Chair Professor

Rua-Huan Tsaih

Rua-Huan Tsaih, Ph.D.
Executive Editor and Distinguished Professor

Department of Management Information Systems
College of Commerce
National Chengchi University
Taipei, Taiwan
Fall 2017

MIS Review 22(1/2), September (2016)/March (2017), 1-20.

Hidden Topic Analysis for Personalized Document Recommendation

Te-Min Chang[1], Wen-Feng Hsiao[2], Ming-Fu Hsu[3]

[1] *Department of Information Management, National Sun Yat-sen University*

[2] *Department of Information Management, National Pingtung University*

[3] *English Program of Global Business, Chinese Culture University*

ABSTRACT: *Collaborative filtering (CF) is a common technique used by most recommender systems for reducing information overload or finding products to purchase. The limitation for applying of CF on document recommendation lies in that CF does not consider the semantics of documents, but rather collects preferential information from many users. In literature, document recommendation tasks rely more on content-based filtering (CBF) simply because document contents are handy for analysis. However, CBF methods alone have a natural limitation in the feature selection to represent the items and in serendipitous recommendations beyond items in the profile.*

The objective of this research is thus to propose hybrid filtering approaches for document recommendation system to overcome the shortcomings of each method alone. Particularly, latent Dirichlet allocation (LDA) was incorporated to uncover latent semantic structure in the collected document corpus. The hidden topic results can act as the technical bridge between CBF and CF because we can either obtain robust document similarity in CF, or to further explore user profiles in CBF. Two experiments were conducted accordingly. The results showed that our proposed approaches outperform other counterparts on the recommendation performance, which justifies the feasibility and practical applications of the proposed approaches.

KEYWORDS : *Personalized Document Recommendation, Content-Based Filtering, Collaborative Filtering, Hidden Topic Analysis, Latent Dirichlet Allocation.*

1. Introduction

With the explosive growth of information over the Internet, more and more information is disseminating and exchanging through this new channel. The large amount of information, however, results in a challenge for users to find relevant information they are interested in. This is commonly referred to as the information overload problem due to human's limited information processing ability. To alleviate this problem, many researchers have resorted to information retrieval (IR) and information filtering techniques that help practitioners develop various tools such as search engines and recommender systems to facilitate users' online information acquisition and filtration.

Among others, recommendation service has been successfully applied to support users to identify desired information by filtering undesired one. Recommender systems suggest users the relevant information through the analyses of their past preferences or the preferences of like-minded people. Accordingly, two types of filtering techniques in recommender systems have been proposed: content-based filtering (CBF) and collaborative filtering (CF). Content-based filtering techniques compare the new information with an active user's profile of past interest to predict whether he/she whom the

recommendation/prediction is for is interested in the new information, whereas collaborative filtering techniques look for collective preferences from other similar users, and recommend their common interests to the active user.

Personalized document recommendation task is essential in assisting users to locate the preferred information in terms of textual content. In literature, CBF technique is primitively employed because the document contents are handy for analysis. However, CBF easily suffers from over-specialization problem that only recommends contents highly similar to the active users' past preferences, and thus performs worse than CF. Extra information is needed to bridge the CBF and CF in personalized document recommendation tasks to enhance the resultant performance.

Recently, several hidden topic analysis approaches such as probabilistic latent semantic analysis (pLSA) and latent Dirichlet allocation (LDA) have been developed. They serve as dimension reduction approaches to investigating content features by revealing hidden topics of documents from the document corpus. Such information acts perfectly as the technical bridge because we can either explore user profiles in CBF or obtain robust document similarity in CF based on the hidden topic results. Chances are that better document recommendations can be generated with the incorporation of hidden topic information.

The objective of this research is thus to propose hybrid filtering approaches for the task of personalized document recommendation. Particularly, LDA was employed to uncover the semantic structure hidden in the document corpus. The semantic structure can account for the word distributions over the latent topics and for the latent topic distributions over documents. We can utilize LDA results in such ways as exploring user profiles to further understand users' preferences, or enhancing robustness of document similarity measures. It is desired that our proposed hybrid approaches can compensate for traditional CBF or CF alone to yield better recommendation predictions.

The rest of the paper is organized as follows. In Section 2, related work is introduced, which include filtering approaches in recommender systems and hidden topic analysis models in text mining fields. Section 3 presents our proposed approaches. Experimental results and corresponding findings are shown in Section 4 to justify the proposed approaches. Finally, concluding remarks are addressed in Section 5.

2. Related work

In the mid-1990s, recommender systems that provide recommendation service have appeared to be an important research area to help users overcome information overload problem and to provide personalized recommendations (Adomavicius & Tuzhilin, 2005). The purpose of recommender systems is to facilitate the information filtering process by automatically recommending desired information (e.g. books, CDs, videos, movies, and news) through the analysis of our past preferences or the preferences of other individuals who share similar interests.

When it comes to document recommendation service, the major concern is to assist users to acquire the preferred information in terms of textual content. It becomes more and more important especially under the information overload environment like the Internet where there is more than enough information disseminating and circulating around. Circumstances such as assisting researchers to decide which scholarly papers to read, supporting knowledge workers to access task-related documents to perform tasks, or

optimizing the learning environment for the learners in e-learning systems are particularly in need of the assistance of personalized document recommendations.

Due to the textual content nature, early works on document recommendations mostly rely on content-based filtering approach. New documents are recommended by matching their content with the user profile consisting of his/her preferred content features. This allows explanations of which documents to be recommended with distinct content features. For example, Mooney and Roy (2000) developed text classifiers for the task of book recommendation. However, the recommendation performance may not be satisfactory because few serendipitous recommendations beyond those documents in the profile can be generated.

In literature, collaborative filtering approach is proposed to recommend items that are beyond those in the user's profile based on other users' collective preferences. It can be further categorized into two general classes: memory-based and model-based (Adomavicius & Tuzhilin, 2005). Acknowledging the fact that documents contain latent topics, several research works attempt to address document recommendations via model-based CF approaches (Cleger-Tamayo, Fernández-Luna, & Huete, 2012; Luostarinen & Kohonen, 2013). The model-based approaches build a model (or classifier) based on the collected user-item ratings, terms, and categories. Models can be in the form of clusters, k-nearest neighbor algorithms, Bayesian networks, or probabilistic relations. Such approaches can be efficient once the model is established; however, they may not fit under the no preferential rating (read or unread) situation as they reduce to one-class classification problems, which are more difficult to confront.

However, only a few works address document recommendations via the collaborative filtering approach. The reason may lie in the rare rating information provided by the readers compared to the abundant number of words obviously embedded in documents. As a counterexample, Amazon.com employed item-based collaborative filtering to recommend books (Linden, Smith, & York, 2003) since it is relatively easy for this company to collect adequate data as the world's largest online retailer. Hess, Stein, and Schlieder (2006) proposed to integrate a document reference network and a trust network for document recommendation. The trust network is derived from trust-based collaborating filtering that utilizes extra trust information to complement original rating information.

To bridge the techniques of CBF and CF in personalized document recommendation tasks, we need extra information to proceed. Recently, hidden topic analysis such as latent Dirichlet allocation (LDA) (Blei, Ng, & Jordan, 2003) has been proposed to uncover the hidden topics of the semantic structure in the document corpus. LDA has been applied in several fields such as text segmentation (Misra et al., 2011), tag recommendation (Krestel, Fankhauser, & Nejdl, 2009), automated essay grading (Kakkonen et al., 2008), fraud detection in telecommunications (Xing & Girolami, 2007), and Web spam classification (Bíró et al., 2009). LDA can be applied in document recommendation as well because the discovered hidden topics serve perfectly as the link to incorporate CBF and CF to enhance the recommendation performance.

3. Proposed approaches

As stated, the objective of this research is to propose hybrid filtering approaches for personalized document recommendation. Latent Dirichlet allocation (LDA) is employed to uncover the semantic structure hidden in the documents. After LDA model is established, we propose two different approaches

to incorporating the LDA results into recommendation. The first one is to apply them directly into item-based CF similarity computation to facilitate the CF prediction process, i.e., the document similarity is measured by the latent topic distributions of two documents. The second approach is to explore user profiles where latent topics of each profile are revealed by applying the LDA results. The former approach is hereinafter referred to as semantic-based collaborative filtering (SBCF), and the latter approach as collaborative-based profile filtering (CBPF). Figure 1 shows the steps of both approaches, which are described in details in the following sections.

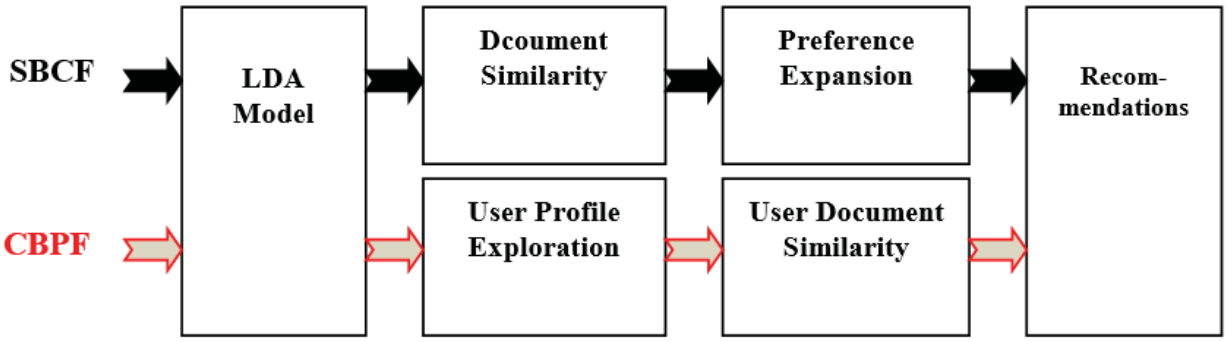


Figure 1 Steps of proposed approaches

3.1 LDA modeling

The first step of both SBCF and CBPF approaches is to build the LDA model from the collected documents which users have seen or liked. LDA is a generative probabilistic model for documents. It assumes that each document is composed of a mixture of latent topics that follow a multinominal distribution with parameters generated by Dirichlet distribution, and each latent topic is composed of a mixture of words that follow a multinominal distribution with parameters generated by Dirichlet distribution. Figure 2 shows the generative model of LDA where M denotes the number of documents (d), N denotes the number of words (w) in a document, K is the number of topics (z), α is the Dirichlet parameter specifying the document-topic distributions θ, and β is the Dirichlet parameter specifying the topic-word distribution φ.

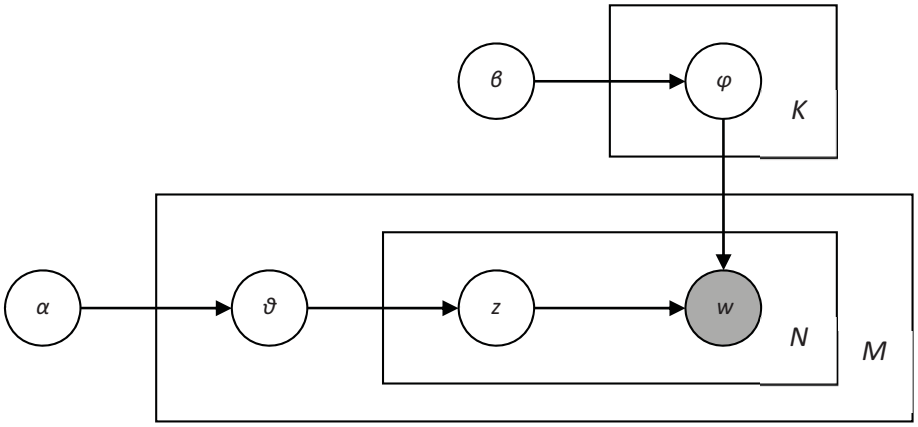


Figure 2 LDA generative model

The major task in LDA modeling is to estimate the two random variables, θ and φ, along with their associated parameters α and β, respectively, from the collected unlabeled corpus. It is shown in literature that variational EM (Expected Maximization) (Blei, Ng, & Jordan, 2003) and Gibbs sampling (Griffiths & Steyvers, 2004) are two common approaches that can be applied for the estimation. Nevertheless, most of research works focus on Gibbs sampling since its performance is comparable to variational EM but faster in convergence and better tolerant to local optima (Darling, 2011).

Accordingly, in this study, we employ the Gibbs sampling to estimate parameters of LDA which iterates multiple times over each word w to sample a new topic k for the word based on the probability $p(z_i = k|w, z_{-i})$ as follows:

$$p(z_i = k|w,\ z_{-i}\ ) \propto (n_{d,k} + \alpha K)\ \frac{n_{k,\ w} + \beta w}{\sum_{w'} n_{k,\ w'} + \beta w} \tag{1}$$

where $n_{k,w}$ maintains a count on topic-word assignments, $n_{d,k}$ counts the document-topic assignments, z_{-i} stands for all topic-word and document-topic assignments except the current assignment z_i for word w, W is the total vocabulary in the document collection, and α and β are the parameters for the Dirichlet priors, serving as smoothing parameters for the counts. Through the counts of posterior probabilities in eq. (1), parameters θ and φ can be obtained as follows:

$$\theta_{d,k} = \frac{n_{d,k} + \alpha K}{\sum_{k'} n_{k',\ w} + \alpha K} \tag{2}$$

$$\varphi_{k,w} = \frac{n_{k,\ w} + \beta W}{\sum_{w'} n_{k,\ w'} + \beta W} \tag{3}$$

3.2 Semantic–Based Collaborative Filtering (SBCF)

After the LDA modeling step, SBCF utilizes the hidden topic results to serve as a basis of document similarity measures. Such similarity results can be easily applied in the usual item-based CF prediction process. Consequently, SBCF goes through steps of measuring document similarity, expanding active user's preferences, and making the final recommendation predictions. These steps are further discussed in the following.

3.2.1 Measuring document similarity

This step is to use LDA results to find out the similarity between documents in order to facilitate item-based CF prediction. The estimated θ denotes the latent topic distribution of each document. It is viewed as a matrix of documents by topics, and can be applied to calculate the similarity between documents. Here we apply the cosine similarity as the similarity measure between any two documents and we can obtain a document similarity matrix S as a result.

For example, assume that we collect three documents and fix the number of the latent topics to be three. The LDA results for θ are further assumed as those shown in Table 1. From Table 1, the first document (D_1) is distributed over the three topics (T_1, T_2, and T_3) with probabilities of 0.2, 0.5, and 0.3, respectively. Documents D_2 and D_3 can be interpreted in a similar manner. Therefore, we can apply the similarity measure between any two documents and get a similarity matrix S. Table 2 shows the document similarity results using the data in Table 1. Note that in this simple example, D_1 is highly similar to D_2 with a similarity degree of 0.925 because both emphasize T_2 more but T_1 and T_3 less. While D_2 and D_3 do not share a lot in common with a similarity degree of 0.309 because D_3 emphasize T_1 more but T_2 less.

Table 1 Estimated θ of the LDA model

	T_1	T_2	T_3
D_1	0.2	0.5	0.3
D_2	0.1	0.75	0.15
D_3	0.7	0.1	0.2

Table 2 The document similarity matrix S

	D_1	D_2	D_3
D_1	1	0.925	0.552
D_2	0.925	1	0.309
D_3	0.552	0.309	1

3.2.2 Expanding active user's preferences

In this step, we desire to expand active user's preferences based on the document similarity result to predict his/her interests toward unseen documents. Assume that we are given users' reading records of documents as a matrix R, which is the rating matrix employed in CF in the typical sense. In our study, the element values of R are either "1", denoting the user has seen and liked this document, or "–", denoting the user has not seen this document yet. We then look at the set of documents an active user has seen and determine how similar they are to other documents the active user has not seen using the similarity matrix S from the previous step. In this regard, the ratings of unseen documents for the active user can be obtained and they serve to indicate the preference degrees of the active user toward the unseen documents.

For item-based CF, the predicted rating $P_{a,i}$ for a novel document i, with respect to an active user a, is based upon the weighted average of ratings from all other documents that have been rated by the active user a. In our study, however, we assume no preferential degrees in R but only "1" or "–". This is somehow limited by the data collection because in personalized document recommendation, we usually obtain what documents users have read instead of the degrees to which documents users have read and liked. This implies that traditional prediction formula for $P_{a,i}$ cannot be directly applied. We therefore modify the formula into the sum of similarity degrees between all read documents n and the unseen document i, as follows

$$P_{a,i} = \sum_{n \in N_a} w_{i,n} \tag{4}$$

where N_a is the document set that user a has read and $w_{i,n}$ is the similarity degree between documents i and n from the similarity matrix S.

Again, we take a simple example to illustrate the above processes. Assume that we obtain the users' reading records R as in

Table 3. The active user 2 has seen documents D_1 and D_2 but not D_3. To infer how user may like D_3, we simply sum up the similarity degrees of D_1 to D_3, and D_2 to D_3, which are 0.552 and 0.309, respectively from the similarity matrix S in Table 2. Then

$$P_{2,3} = 0.552 + 0.309 = 0.861 \tag{5}$$

is the expanded preference for user 2 toward unseen document D_3. The final results of the expanded preferences in this example are listed in Table 4.

Table 3 Users' reading records of matrix R

	D_1	D_2	D_3
U_1	—	—	1
U_2	1	1	—
U_3	—	1	—

Table 4 Expanded users' preferences of documents

	D_1	D_2	D_3
U_1	0.552	0.309	1
U_2	1	1	0.861
U_3	0.925	1	0.309

3.2.3 Making top-N recommendations

Finally, SBCF performs the final step of recommendation predictions based on the predicted ratings for unseen documents based on equation (4). We simply rank the ratings in a descending order and select the first N documents to generate the top-N recommendation list for each active user.

3.3 Collaborative-Based Profile Filtering (CBPF)

After the LDA modeling step, CBPF utilizes the hidden topic results to explore user profiles. The similarity between an unseen document and the user profile can then be easily measured over the latent topic distributions. Consequently, CBPF goes through steps of exploring user profiles, measuring user-document similarity, and making the final recommendation predictions. These steps are further discussed in the following.

3.3.1 Exploring user profiles

This step is to apply the LDA results to explore each user's preference over the latent topics in his/her profile. By transiting the user-document relationship (from the rating matrix R) and the document-topic relationship (the estimated θ random variables in LDA model) with average operation into the user-topic relationship, the topic distribution in a user profile can be easily inferred as follows:

$$T = a \times R \times \theta \qquad (6)$$

where a is the normalization factor for each row in T and "–" in R can be deemed as 0 under the matrix multiplication operation.

Assume again the LDA results for θ are shown in Table 1, and the users' reading records of documents of matrix R are shown in

Table 3. Note that each row of R serves as each user's reading profile. We average the topic distributions of documents in the user profile to obtain the inferred latent topic distribution for this profile. The results are shown as in Table 5. The inferred topic distribution indicates how preferable the topics interest the user.

Table 5 Topic preferences of users

	T_1	T_2	T_3
U_1	0.7	0.1	0.2
U_2	0.15	0.625	0.225
U_3	0.1	0.75	0.15

3.3.2 Measuring user-document similarity

In this step, we desire to calculate the similarity between each unseen document and the user profile in terms of latent topic distributions based on the cosine similarity measure. After the similarity computation is performed over all documents and all users, we can get a similarity matrix I as

$$I = T \times \theta^T \tag{7}$$

where θ^T is the transport matrix of θ. In fact, the similarity results for unseen documents are simply their predicted ratings.

Again, take the previous simple example. Table 1 shows how the three documents are distributed over the latent topics and Table 5 shows how the three user profiles are distributed over the latent topics. We then apply the cosine similarity measure between users and documents. Table 6 shows the similarity results for the similarity matrix I. Note that in this simple example, U_1 would fairly prefer unseen document D_1 (with 0.5519 similarity degree) and U_3 would highly prefer unseen document D_1 (with 0.9253 similarity degree).

Table 6 Similarity between documents and user profiles

	D_1	D_2	D_3
U_1	0.5519	0.3087	1.0
U_2	1.0	1.0	0.4246
U_3	0.9253	1.0	0.3087

3.3.3 Making Top-N recommendations

Finally, with the results of matrix I, CBPF can proceed in the final step to sort the ratings for the unseen documents in the descending order and select the first N documents that are of top-N predicted ratings to generate the recommendation list for each active user.

4. Experiments and results

In this section, we conduct two experiments to examine the performance of our proposed hybrid filtering approaches. In our experiments, we collect the dataset from CiteULike (http://www.citeulike.org/), which is commonly applied in document recommendation and tag recommendation. CiteULike is a website that assists users to store, organize, and share scholarly papers. Users can annotate the scholarly papers they are interested in with tags (bookmarks). Therefore, the information provided in CiteULike fits appropriately the domain of document recommendations.

We utilize the bookmark data collected from January 1st 2012 to April 30th 2013 as our experimental data. Essentially, we only extract the document abstracts because they ease the computational efforts of LDA analysis, which may be considered as a restriction of LDA applications. Furthermore, we filter out users who read less than 20 documents in their personal profile since it is more difficult and unreliable to predict these cold start users. We also filter out those documents that occur only once during the time period because the cold start items do not contribute significantly in the analysis either. We therefore obtain a dataset, called CUL, consisting of 495 users, 13,029 documents, and 36,466 bookmarks.

Our study adopts *Precision*, *Recall,* and *MAP* that are commonly applied in information retrieval fields to measure the recommender performance. *Precision* is the fraction of recommended items that are relevant. It is defined as the number of hits (i.e. the number of documents in the test set that also appears in the top-N recommended documents) divided by the number of all recommended documents, as defined by:

$$Precision = Number\ of\ hits\ /\ N \tag{8}$$

where N is the number of recommended documents. Higher Precision values indicate more preferable documents are retrieved for the users.

On the other hand, *Recall* is the fraction of relevant instances that are retrieved. It is defined as the number of hits divided by the number of documents in the test set, as defined by:

$$Recall = Number\ of\ hits\ /\ N_{a,t} \tag{9}$$

where $N_{a,t}$ is the number of relevant documents (in our study, it's the number of documents that user has read in the test set).

Finally, average precision (*AP*) is used for evaluation systems that return a ranked list of documents. It considers the precision scores at each ranked position of the returned documents in the list. It is defined as

$$AP = \frac{\sum_i Precision@i \times corr_i}{N_{a,t}} \tag{10}$$

where $Precision@i$ is the precision at rank i and $corr_i$=1 if the document at position i is relevant, otherwise $corr_i$ = 0. $N_{a,t}$ is the number of documents that user has read in the test set. *MAP* is the mean of average precision scores over all test users.

The evaluation scheme used in our approach is the 10-fold cross-validation where the data are randomly divided into 10 equal-sized subsets with respect to the users. Each time, nine of the subsets are prepared for the training and the remaining one subset is prepared for the test. However, the actual training data contain both the 9 subsets and 50% of the remaining subset, randomly selected with respect to each user (as shown in the shaded area of Figure 3). Then the rest withheld 50% of the remaining subset is the

test data (as shown in the blank area of Figure 3) to evaluate the performance. For each user in the remaining subset, we generate a top-N recommended list of documents using the training data and measure the performance for the test data. This procedure is repeated 9 times and the final performance is averaged over the 10 folds to obtain robust results.

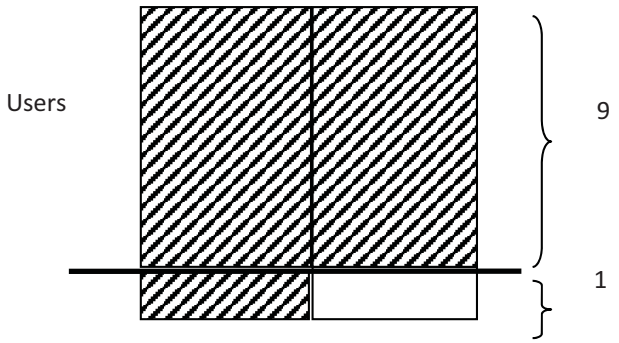


Figure 3 The schematic diagram of evaluation scheme

4.1 Experiment I

The objective of Experiment I is to set up parameters employed in SBCF and CBPF, or more precisely, in LDA model, which includes Dirichelt hyper-parameters α and β, and the number of topics K. Parameter α denotes the Dirichlet distribution parameter over the multinomial distribution of latent topics for each document, and parameter β denotes the Dirichlet distribution parameter over the multinomial distribution of words for each topic. They also serve as smoothing parameters for the counts in Gibbs sampling. In literature, some guidance is provided for these two parameters as which suggested that $\beta = 0.1$ and $\alpha = 50/K$ (Griffiths & Steyvers, 2004). We therefore adopt this setting in our experiment.

The more difficult setting is the number of latent topics, K. It usually varies in different situations such as the selected dataset and its associated size. We therefore select a subset out of the CUL dataset to estimate this parameter. This subset consists of 195 users, 9,616 documents, and 15,260 bookmarks. Once parameter K is estimated, it will be applied to the whole CUL dataset, as shown in the next experiment.

Blei, Ng, and Jordan (2003) proposed a perplexity measure that is commonly applied in language modeling to evaluate the predictive power of the model. The lower the perplexity, the better performance the trained model will be. Therefore, by varying the number of latent topics, we can observe the trend of the perplexity measure and setup the number of latent topics when the trend reaches its minimum.

In our experiment, we use Stanford Topic Modeling Toolbox which was developed by the Stanford NLP group to build the LDA model and measure the perplexity. The result is shown in Figure 4 that illustrates the tendency of perplexity with different number of latent topics. From this result, we do observe a U-shaped curve that reaches its minimum around 80 latent topics.

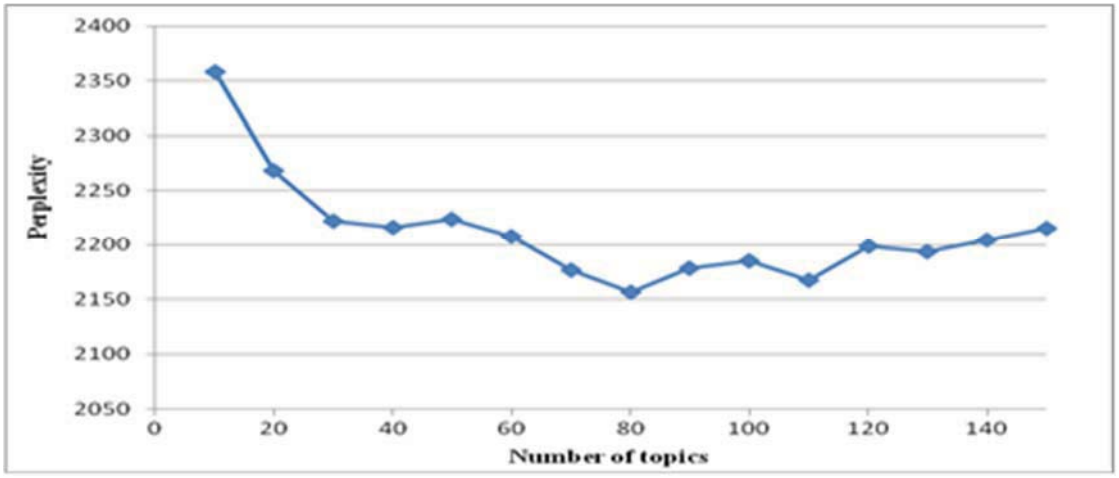


Figure 4 Perplexity of LDA model

However, as mentioned in literature (e.g. Asuncion et al, 2009), perplexity is not always a reliable measure to determine the number of latent topics. Therefore, in our study, we alternatively choose to determine *K* by trial and error that varies from 10 to 300, in increment of 10 each time. The performance results are shown in Figure 5 and Figure 6 for SBCF, and Figure 7 and Figure 8 for CBPF, respectively. They revealed the recommendation performance with different number of recommend documents N and different number of topics.

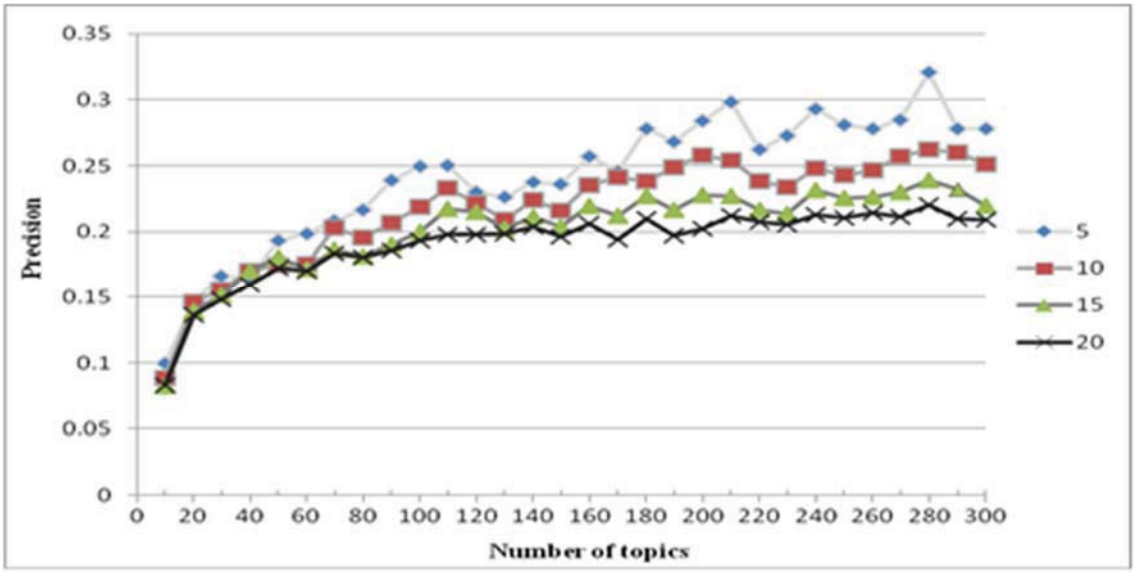


Figure 5 Precision performance of SBCF

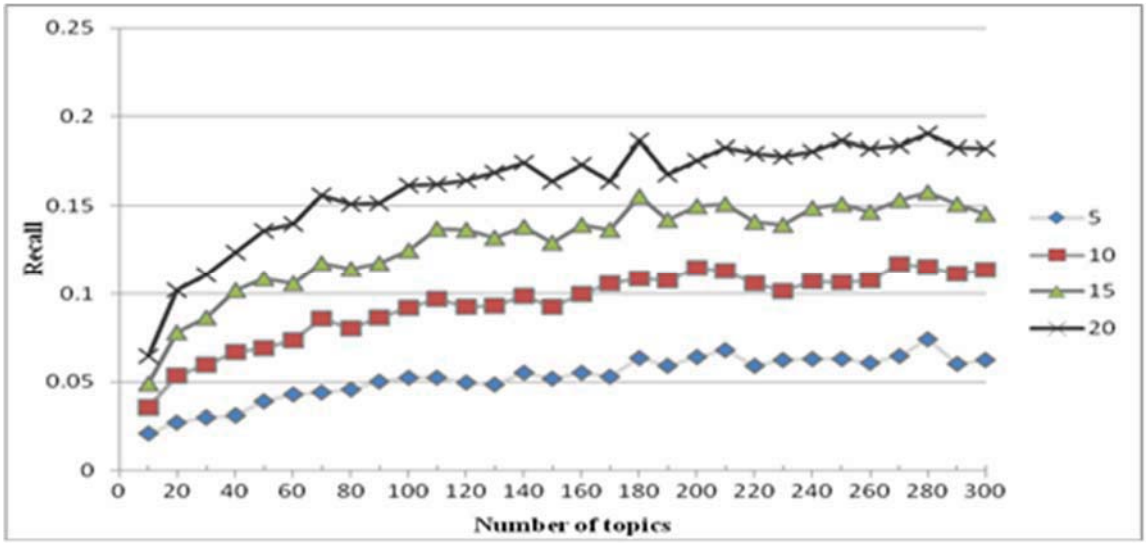


Figure 6 Recall performance of SBCF

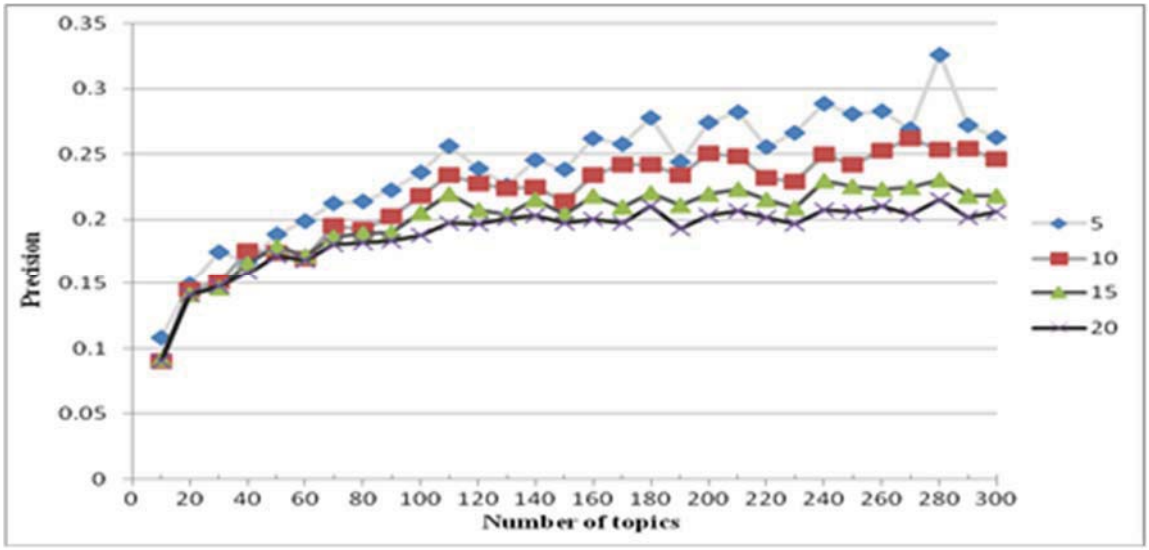


Figure 7 Precision performance of CBPF

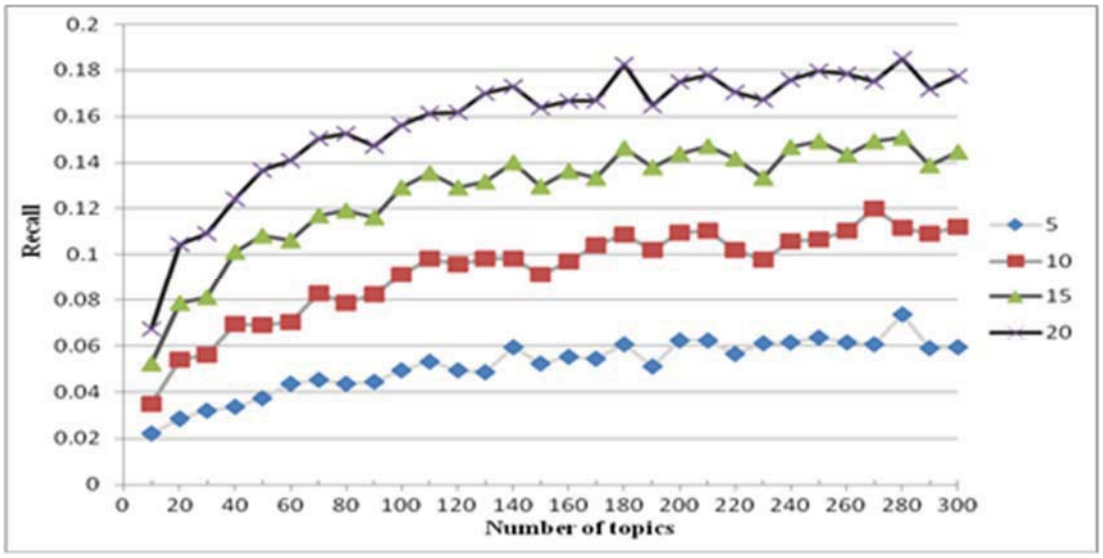


Figure 8 Recall performance of CBPF

From the above figures, apparently the best performance did not occur with the number of latent topics being 80. Instead, the performance fluctuated upwards until it reached the maximum around the number of latent topics of 280. Since latent topics serve as the similarity computation basis for document-to-document similarity in SBCF and document-to-profile similarity in CBPF, both approaches will exhibit its feasibility on recommendation prediction with sufficient (not too few) and non-redundant (not too many) latent topics. To summarize, we set up the parameters employed in the experiments as $\beta = 0.1$, $\alpha = 50/Z$, and the number of latent topics = 280 for both SBCF and CBPF.

4.2 Experiment II

In this experiment, we desire to examine the performance of SBCF and CBPF using the CUL dataset. In addition, we compare their performance with three other approaches: content-based filtering (TFIDF), user-based CF (UBCF) and item-based CF (IBCF) as baselines. For TFIDF, we extract 1000 features with the most TFIDF weights to form the document vector bases. Documents in a user profile will then be aggregated into a profiling document vector for that specific user. We compare a novel document vector with the profiling vector to determine their similarity based on which novel documents are ranked. For UBCF, we do not restrict the neighbor size. The user similarity is determined as the intersection counts of "1"s between two user ratings divided by the union counts of "1"s between them. Likewise, the item similarity in IBCF is determined as the intersection counts of "1"s between two item ratings divided by the union counts of "1"s between them.

Figure 9, Figure 10, and Figure 11 show the performance comparison results of *Precision*, *Recall* and *MAP*, respectively, where N denotes the top-N ratings in the recommendation list.

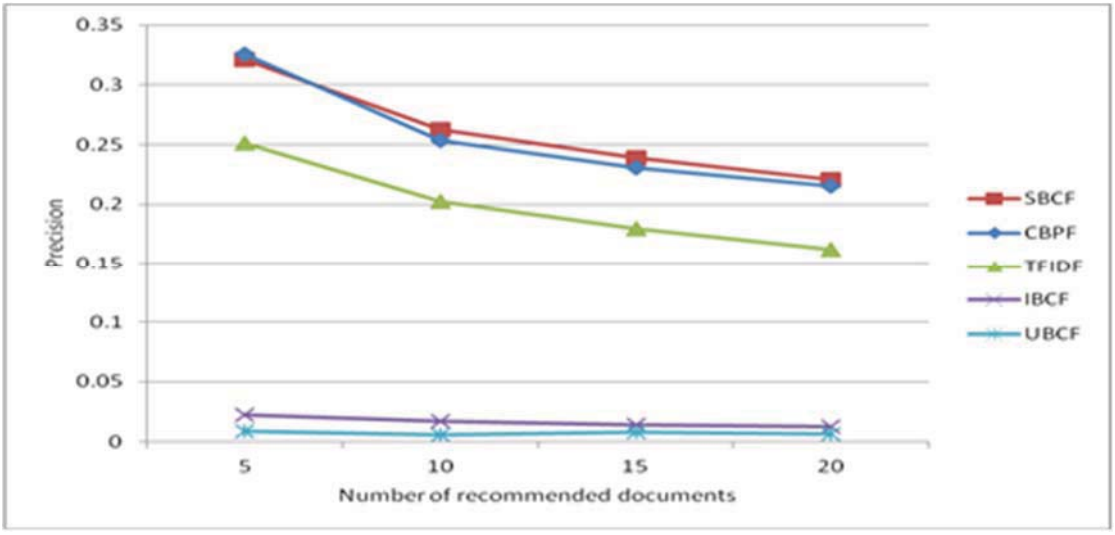


Figure 9 Comparison of Precision performance

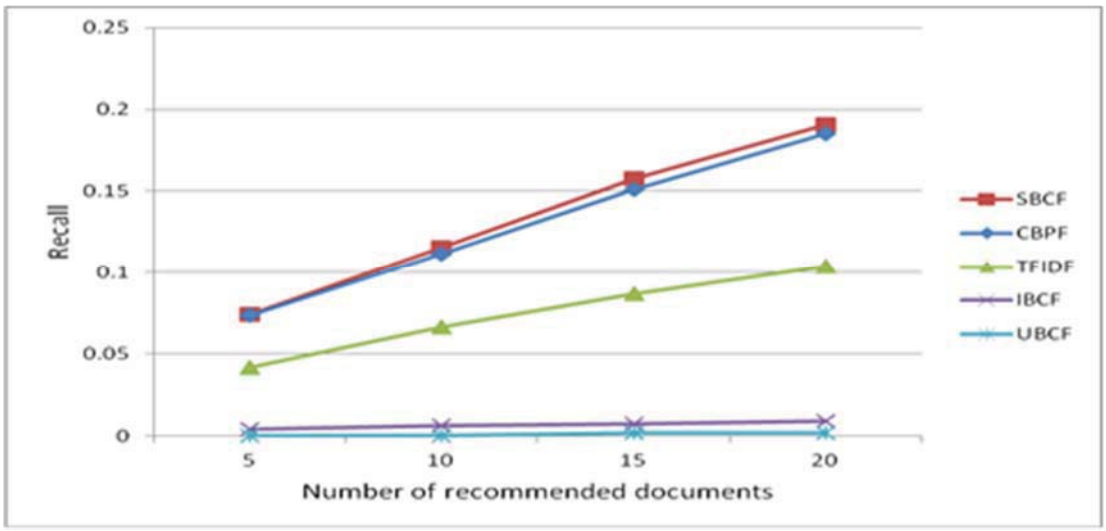


Figure 10 Comparison of Recall performance

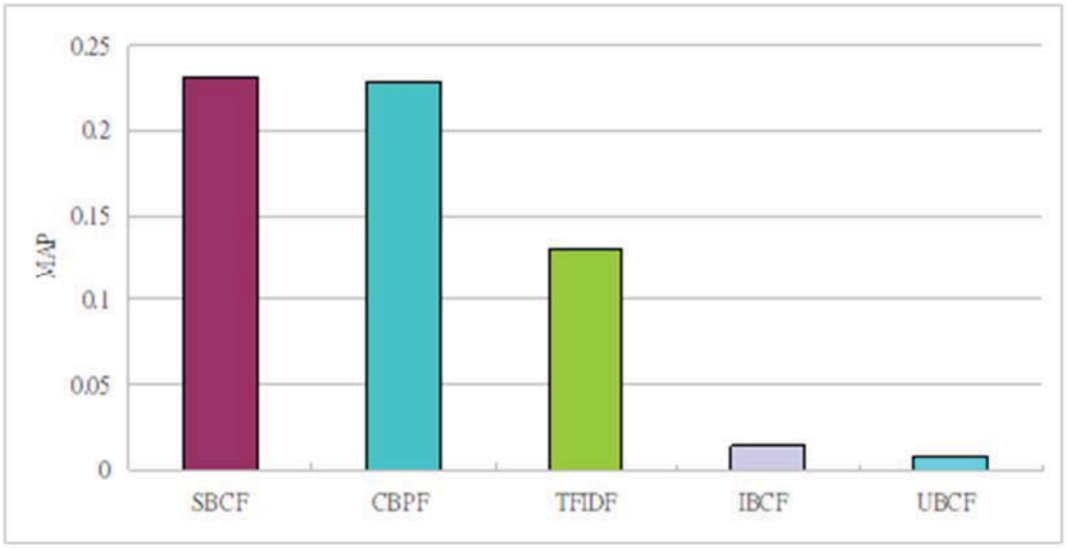


Figure 11 Comparison of MAP performance

From the above figures, we first observe that both UBCF and IBCF perform worst. This is because the traditional CF approaches suffer severely from the sparsity problem in CUL (The sparsity is $1 - 4870 / (86 \times 3201) = 98.23\%$.). This result once again reflects the unreliable predicted recommendation for pure CF approaches if the coverage of the rating matrix is highly sparse.

In contrast, SBCF, CBPF and TFIDF show significantly better performance compared to traditional CF approaches. Personalized document recommendations may well rely on content-based filtering approaches such as TFIDF since the document content is ready for analysis. TFIDF, however, still performs below our expectation since it easily runs into the over-specification problem that cannot expand users' preferences beyond their past profiles.

On the other hand, the results indicate that SBCF does not suffer too much from sparsity because the document similarity calculation is based on the latent topic distributions over documents instead of the document ratings in the rating matrix. The results also indicate that CBPF does not suffer too much from over-specification because the user profiles have been explored based on the latent topic distributions. These outcomes demonstrate the necessity of employing hybrid approaches of CBF and CF for the task of personalized document recommendation, and more importantly, the LDA model incorporated into our proposed approaches exhibits its capability of performing such a task.

Finally, we focus on the comparison between SBCF and CBPF. Both approaches perform seemingly similarly. However, they do differ in their computational efforts. In our experiment, the average computational time is 71.39 seconds under SBCF for all test users in its Step 2 and Step 3, while it is 6.69 seconds under CBPF for all test users in its Step 2 and Step 3. The reason for this substantial difference lies in that SBCF needs to complete the computation of the entire document similarity matrix S before predicting the rating for unseen documents, and its computational time in quadratic proportional to the number of documents. As for CBPF, before predicting the rating for unseen documents, it spends time on investigating the latent topic distributions for user profiles, and its computational time is proportional to the number of documents multiplied by the number of users. In general cases as CUL, the number of users is

far less than the number of documents, and therefore, the time spends on CBPF for the predicted recommendation process will be much less than that on SBCF. To sum up, with similar prediction performance, one may still employ CBPF rather than SBCF due to the time effort for practical considerations.

5. Conclusions

In this research, we propose to utilize the latent Dirichlet allocation (LDA) model to analyze the latent semantic structure among collected documents for personalized document recommendation tasks. With LDA results, latent topic distributions over documents can be uncovered to help either obtain robust document similarity in CF, or explore user profiles in CBF. Two hybrid filtering approaches, SBCF and CBPF, are proposed accordingly in our study.

Two experiments are conducted to examine the performance of our proposed approach. The first experiment is to set up the parameters employed in SBCF and CBPF such as the hyperparameters α and β in the Dirichlet distribution, and the number of latent topics. The second experiment is to compare SBCF and CBPF with traditional content-based filtering (TFIDF), user-based CF (UBCF), and item-based CF (IBCF). The results show that both SBCF and CBPF perform much better than TFIDF, UBCF, and IBCF. The incorporation of the LDA results into the proposed hybrid filtering approaches does enhance the prediction performance significantly.

Finally, the comparison between SBCF and CBPF shows insignificant different performance between them on *Precision*, *Recall* and *MAP*. However, the computational requirement for CBPF is much less than SBCF since SBCF takes time to obtain a full matrix of document similarity before recommendation prediction. Therefore, for practical consideration, one may employ CBPF rather than SBCF to perform the task of personalized document recommendation. To conclude, the experiment results do justify the feasibility of SBCF and CBPF in real applications.

Although the results of our research seem promising, there are some issues that need to be further addressed. First, the "rating matrix" employed in our study does not conform to the usual sense in collaborative filtering because it contains no preferential ratings (such as on the Likert scale) but only "1", indicating the document has been seen and liked by the user. To adapt our proposed approaches into more real situations, we need to collect a more appropriate dataset and examine their feasibility accordingly.

Second, although LDA exhibits its capability of enhancing the document recommendation performance, it has its own limitation of intense computation effort requirements when building the model. Therefore, in current experiments, we only utilize titles and abstracts of documents for analysis and abandon those referred documents without abstracts. However, titles and abstracts only may not be able to reflect the entire semantics in documents and cause the latent topic structure unreliable. A possible resolution to include all referred documents in LDA is to preprocess those documents to reduce their sizes while keep their original semantics. Tasks such as feature selection or text summarization may be considered to apply for the preprocessing step.

Acknowledgment

The author would like to thank Ministry of Science and Technology (MOST) of the Republic of China, Taiwan for financially supporting this work under Contracts No. 101-2410-H-110-005 and No. 105-2410-H-110-033-MY2.

References

Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, *17*(6), 734-749.

Asuncion, A., Welling, M., Smyth, P., & Teh, Y. W. (2009). On smoothing and inference for topic models. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence* (pp. 27-34). AUAI Press.

Bíró, I., Siklósi, D., Szabó, J., & Benczúr, A. A. (2009). Linked latent dirichlet allocation in web spam filtering. In *Proceedings of the 5th International Workshop on Adversarial Information Retrieval on the Web* (pp. 37-40). ACM.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993-1022.

Cleger-Tamayo, S., Fernández-Luna, J. M., & Huete, J. F. (2012). Top-N news recommendations in digital newspapers. *Knowledge-Based Systems*, *27*, 180-189.

Darling, W. M. (2011). A theoretical and practical implementation tutorial on topic modeling and gibbs sampling. In *Proceedings of the 49th annual meeting of the Association for Computational Linguistics: Human Language Technologies* (pp. 642-647).

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National academy of Sciences of the United States of America*, *101*(Suppl 1), 5228-5235.

Hess, C., Stein, K., & Schlieder, C. (2006). Trust-enhanced visibility for personalized document recommendations. In *Proceedings of the 2006 ACM Symposium on Applied Computing* (pp. 1865-1869). ACM.

Kakkonen, T., Myller, N., Sutinen, E., & Timonen, J. (2008). Comparison of dimension reduction methods for automated essay grading. *Journal of Educational Technology & Society*, *11*(3), 275.

Krestel, R., Fankhauser, P., & Nejdl, W. (2009). Latent dirichlet allocation for tag recommendation. In *Proceedings of the Third ACM Conference on Recommender Systems* (pp. 61-68). ACM.

Linden, G., Smith, B., & York, J. (2003). Amazon. com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*, *7*(1), 76-80.

Luostarinen, T., & Kohonen, O. (2013). Using topic models in content-based news recommender systems. In *Proceedings of the 19th Nordic Conference of Computational Linguistics* (pp. 239-251).

Misra, H., Yvon, F., Cappé, O., & Jose, J. (2011). Text segmentation: A topic modeling perspective. *Information Processing & Management*, *47*(4), 528-544.

Mooney, R. J., & Roy, L. (2000). Content-based book recommending using learning for text categorization. In *Proceedings of the Fifth ACM Conference on Digital Libraries* (pp. 195-204). ACM.

Xing, D., & Girolami, M. (2007). Employing latent dirichlet allocation for fraud detection in telecommunications. *Pattern Recognition Letters*, *28*(13), 1727-1734.

About the authors

Te-Min Chang*
Associate Professor
Department of Information Management
National Sun Yat-sen University
Kaohsiung, Taiwan
Email: temin@mail.nsysu.edu.tw

Wen-Feng Hsiao
Associate Professor
Department of Information Management
National Pingtung University
Pingtung, Taiwan
Email: wfhsiao@npic.edu.tw

Ming-Fu Hsu
Associate Professor
English Program of Global Business
Chinese Culture University
Taipei, Taiwan
Email: hsumf0222@gmail.com

*Corresponding author

Te-Min Chang obtained a PhD at Department of Industrial Engineering from Purdue University, USA, in 1996. He is now an Associate Professor at the Department of Information Management, National Sun Yat-sen University, Taiwan. He has published over 20 referred journal papers including Information Sciences, Computers in Human Behaviors, International Journal of Production Research, Annals of Operations Research, and so forth. His current research focuses on text mining, social network analysis, and decision making.

Wen-Feng Hsiao obtained a PhD at Department of Information Management from National Sun Yat-sen University, Taiwan, in 2001. He is now an Associate Professor at the Department of Information Management, National Pingtung University, Pingtung, Taiwan. He has published over 20 referred journal papers including Program: Electronic library and information systems, Journal of Information Science and Engineering, International Journal of Computer Systems Science and Engineering, International Journal of Intelligent Systems Technologies and Applications, and so forth. His current research focuses on text mining, recommendation systems, and decision making.

Ming-Fu Hsu obtained a PhD at Department of International Business Studies from National Chi Nan University, Taiwan. He is now an Assistant Professor at the English Program of Global Business, Chinese Culture University, Taiwan. He has published more than 20 journal papers including Knowledge-Based Systems, Information Sciences, Neural Computing and Applications, Economic Modelling, Connection Science, and so on. His current research focuses on financial data analysis, text mining, social network, and performance measure.

♦

MIS Review 22(1/2), September (2016)/March (2017), 21-40.

Factors Influencing Consumers' Attitudes toward Social Media Marketing

Asad Ahmad, Mohammed Naved Khan
Department of Business Administration, Faculty of Management Studies & Research, Aligarh Muslim University

ABSTRACT: *With the growth of Internet social media has transformed itself into an innovative platform where people from different regions interact and share experiences with each other. People with common interest interact and share their experiences with the other members of the Social Networking Sites (SNS). SNSs have become a center stage for e-commerce with the phenomenal rise in the number of the SNSs users. It has become a platform where different marketers advertise their products and services. This medium has equipped the marketers in reaching their target consumers easily. The researchers in the present study have used an 18 item scale to study the factors determining attitude of the students towards advertisement over SNSs. Usefulness, Reliability and Word of Mouth Quality emerged to be the critical factors determining the attitude of the SNSs users towards the advertisements.*

KEYWORDS : *Internet, Social Media, Social Networking Sites (SNS), Social Media Marketing, Advertisements.*

1. Introduction

The rapid growth of Internet in the last decade has made it a vital part of each and every soul. The Internet users' base worldwide has grown to 3.42 billion in 2016 from 2.92 billion in 2014 (Internet Live Stats, 2016). Internet is being used for various activities like searching product features, compare prices, select and place order and make payments (Senthil, Prabhu, & Bhuvaneswari, 2013; Sinha, 2010). The fast and flexible communication competencies of the Internet has made it a platform where people communicate and exchange information with others (Ahmad, Rahman, & Khan, 2017; Ahmad, Rahman, & Khan, 2016; Ahmad & Khan, 2015). Internet has evolved itself into an interactive and collaborative from that of passive readings (Sharma, 2008). The interactive quality of the internet has given its' users a new way to interact. Online Social Networking Sites (SNS) are the digital networks in which there occurs an intrinsic connection with other members present over the site (Wellman & Gulia, 1999). Social media has proved itself to be the most significant medium which has extended its root to different segments of the world population (Corbett 2009; Barnes & Mattson, 2008; Bernoff, Pflaum, & Bowen 2008). Traditional channels like television, radio, newspaper etc. delivers single-directional communication but with the advent of Internet and social media the users are able to get two-way conversations (Mayfield, 2008). Among the different types of social media SNSs have been considered to be the most important for the digital media revolution (Vogt & Knapman, 2008). It has also been reported that e-mail has been replaced by SNSs to be the most popular online activity and the primary source from where the Internet users seek information (Fuscaldo 2011; Albanesius 2010).

The growing popularity of the social networks and the ever changing lifestyle of the consumers have brought a revolution in the interaction of the consumers with different brands. The interactive property of the social media have provided the consumers with the power to shape brand and product images through word of mouth in the online connections (Muntinga, Moorman, & Smit 2011). It has been found that consumers now search brand and product related information over social network rather than the corporate websites (Dei Worldwide, 2008). With the increasing popularity of social networking sites it has grabbed the attention of the marketers as a center stage in e-commerce and they are thinking of way and means of using these sites as an advertising medium (Bausch & Han, 2006). These networks has also been used by the marketers to reach out to their target population where they are able to build relationships with their consumers on a more personal level (Kelly, Kerr, & Drennan, 2010). The marketers are trying hard to use these mediums to engage their consumers (Avery et al. 2010). The marketers must keep a balance while advertising through SNSs as many users may not be fine with the unwarranted advertisements on their social mediums (Mesure & Griggs, 2007; Nutley, 2007).

2. Literature review

The growth of social media has brought a communication revolution which helps in creating social relationships platforms over the Internet. It has also helped in expanding the social ties which have increased the psychological well-being among the Internet users worldwide (Thoren et al., 2013; Nabi, Prestin, & So, 2013). In the present online landscape consumers have been turning to the Internet and social media to express their opinion and to exchange information, which have made these mediums an ideal tool for the advertisers to build relationships with their target consumers (Hair, Clark, & Shapiro, 2010). Online consumers extract the product related information from various sources, especially through the product reviews on the social networking sites (Clemons, 2009). Before actual buying of products and services, consumers continuously search for the product reviews (Akar & Topçu, 2011). It has been found that sixty percent of the online consumers do believe on the product reviews by different consumers (Blackshaw & Nazzaro, 2006). The e-wom is taken more seriously than the marketing messages (Akar & Topçu, 2011). The online consumers who use social networking sites have been found to be vital as they are supposed to be active and effective, they share their experiences with other consumers through the social media (Blackshaw & Nazzaro, 2006). If the Internet users use blogs and finds it important then it is likely that they will use them again. The intention to use the blogs again is also impacted by the level of satisfaction of the users (Shiau & Luo, 2010). The Social media users' shops through the social networking sites when they find the services over the social sites are useful as well as easy to use (Cha, 2009).

Social media marketing is the use of different forms of social media to promote the products and services of a company. These kind of online marketing activities generally completes the traditional Internet promotional strategies like e-mail and Internet advertising campaigns (Barefoot & Szabo, 2010). Social media transforms the consumers into marketers and advertisers, who in turn can create positive or negative impact for the products and services offered by the company (Roberts & Kraynak, 2008). Through social media marketing individuals are capable of presenting their products and services to a large community and to get their feedback through social networking sites which is not possible in general advertisements (Weinberg, 2009). According to Ontario (2008) social networking sites helps the marketers to exchange thoughts and information related to product and services. Viral advertising has been defined as "unpaid peer-to-peer communication of provocative content originating from an identified sponsor using the

Internet to persuade or influence an audience to pass along the content to others" (Porter & Golan, 2006). In these form of advertising campaigns the advertisers first send the message to their potential consumers who further share the information with other potential consumers (Southgate, Westoby, & Page, 2010; Dobele et al., 2007). It has gained tremendous popularity with the advent of social media. The way the consumers' responds to advertisements have dramatically changed because of the Internet and social media (Marken, 2007). Facebook and other social media have helped the marketers in transforming their target consumers to become message senders when the users share the advertising message with their group or friends (IAB, 2009).

The purchasing decision of a large chunk of Internet users are affected by the Social Networking Sites hence the use of social media as a marketing medium has become an effective marketing strategy (Miller & Lammas, 2010). It is very important for the e-retailers to understand the consumers and their activities over social media to engage them (Goh, Heng, & Lin, 2013; Rapp et al., 2013).

The number of researches examining the impact of word of mouth on product sales and marketing strategies, consumer product reviews and its' usefulness in consumer decision making and the usefulness of consumer reviews in sales forecasting have increased with the growing number of social network users (Chen, Wang, & Xie, 2011; Dhar & Chang 2009; Chen & Xie 2008; Sen & Lerman 2007; Dellarocas, Zhang, & Awad 2007; Chevalier & Mayzlin 2006; Liu 2006; Mayzlin 2006; Smith, Menom, & Sivakumar 2005; Godes & Mayzlin, 2004). Youth are considered the most ideal group in online setting (Bolton et al., 2013). Thus it is very important to study the Internet and social networking patterns of this group as the behaviors of this group is likely to differ with different contexts. There are literatures demonstrating the variance in the behaviors of males and females with respect to e-commerce but generation Y is yet to be tested (Ruane & Wallace, 2013; Yeh, Hsiao, & Yang, 2012; Garbarino & Strahilevitz, 2004; Rodgers & Harris, 2003). Generation Y is supposed to be active users of social media where they search and share product related information (Bolton et al., 2013). This group have lived their whole life in the digital environment and are profoundly affected by information technology (Bennett, Maton, & Kervin, 2008).

This study aims to study the factors which influence the attitude of the social media users towards the advertisements over the social media. The researchers in the present study have proposed a research model (Figure 1) with *Perceived Usefulness, Reliability and Word of Mouth Quality* to better explain the *Attitude* towards social media advertisements.

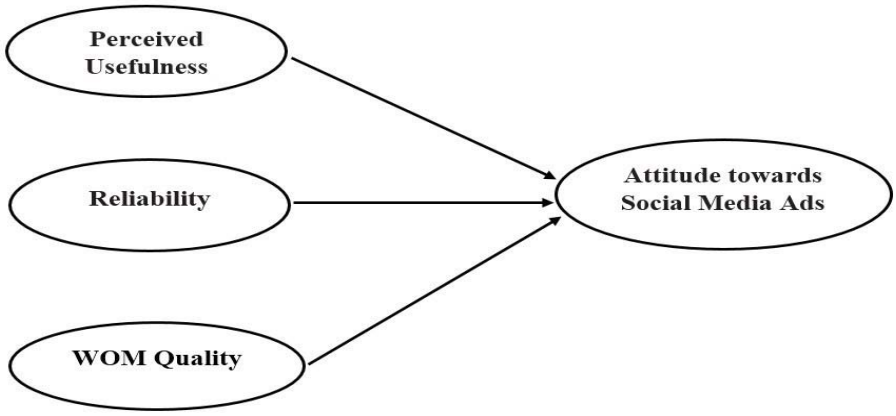


Figure 1 Research Framework

2.1 Hypothesis

The following hypotheses were proposed with respect to the above literature and the proposed research framework:

H1: Perceived Usefulness has a positive effect on Attitude towards social media advertisements.

H2: Reliability has a positive effect on Attitude towards social media advertisements.

H3: Word of mouth quality has a positive effect on Attitude towards social media advertisements.

2.2 Methodology

2.2.1 Survey instrument

The researchers have adapted an eighteen item research instrument that comprised of 4 items on Attitude (adapted from Khare & Rakesh, 2011); 4 items on Perceived Usefulness (adapted from Childers et al.,2001) ; 7 items on Reliability (adapted from Kim, Chung, & Lee, 2011); and 3 items on Word of Mouth Quality (adapted from Awad & Ragowsky, 2008). The respondents of this study were heavy Internet users having accounts in Social Networking Sites (SNSs) for the last one year.

2.2.2 The sample

India is a country of youths where around 50% of the population is below the age of 25 years (Heslop, 2014). And also the majority of the Internet users in India are young i.e. 15-24 years of age (Statista, 2016). The sample of the study comprised secondary and senior secondary school and college going students of a government funded premier central university of India. Middle class strata of the society, who are considered as the representative of the masses (Ahmad & Khan, 2017; Shabnam, 2012) are concentrated in these kind of institutions (Heslop, 2014). Thus, the sample may be deemed representative of the entire student population.

2.2.3 Scale development

The variable items were modified and rephrased by the researchers keeping in mind the profile of student respondents. As per the suggestions of Liao et al. (2011) a pilot study was conducted to check the appropriateness of the scale in the Indian context. Responses were generated from 40 social media users for the pilot study. On the basis of the responses various items were re-phrased. To check the appropriateness of the scale on Indian setting, it was analyzed using Exploratory Factor Analysis (EFA). Items which cross loaded or had loadings less than 0.5 were dropped (Khan & Adil, 2013; Metin et al., 2012; Büyüköztürk, 2003; Yoo & Donthu, 2001; Hair et al., 1998, Kline, 1994; Anderson & Gerbing, 1988). The analysis resulted in a shorter 13-item scale comprising four variables (*Attitude, Perceived Usefulness, Reliability and Word of Mouth Quality*).

2.2.4 Data collection

To collect data, the researchers employed convenient sampling technique. The refined questionnaires were administered to secondary, senior secondary, graduates and post-graduate students. The researchers personally administered the questionnaires to clarify doubts which enhances the quality of the responses (Dornyei & Taguchi, 2010). A total of 195 questionnaires were distributed of which 160 questionnaires were returned. Out of the 160 returned questionnaires 140 were usable responses. Table 1 provides the summary of the demographic profile of the respondents.

Table 1 Demographic Profile of Respondents

	Frequency
Qualification	
Secondary & Senior Secondary	75
Graduates & Post graduate	65
Age	
Less than 17 Years	67
Above 17 Years	73
Gender	
Male	80
Female	60
See ads over social media	
Yes	133
No	07

\

3. Results

3.1 Exploratory Factor Analysis (EFA)

EFA was performed to determine the Principal Component Analysis (PCA) with varimax rotation and Kaiser normalization as the factor extraction method using SPSS 20. The items indicated a practical level of variance as the KMO measurement of sampling adequacy value was found to be 0.728 (KMO > .6), in addition, the Bartlett's test of Sphericity value (Chisquare = 1654.781, $p < 0.005$) was found to be significant (Herington & Weaven, 2007; Hair, Anderson, Tatham & William 1998), which proved that the sample size of 140 was significant for analysis. Items with value .5 or more are acceptable (Khan & Adil, 2013; Metin et al., 2012; Büyüköztürk et al., 2004; Hair et al., 1998). Table 2 shows the loadings of the various items which are found to be acceptable i.e. more than 0.6. The Cronbach's Alpha coefficient of Factor "Attitude" (3 items), "Perceived Usefulness" (4 items) and "Reliability" (3 items) and "Word of Mouth Quality" was found to be 0.675, 0.712, 0.693 and 0.670 respectively. The Cronbach's Alpha coefficient of all the four variables were within acceptable range (Khan & Adil, 2013; Kerlinger & Lee, 2000; Hair et al., 1998).

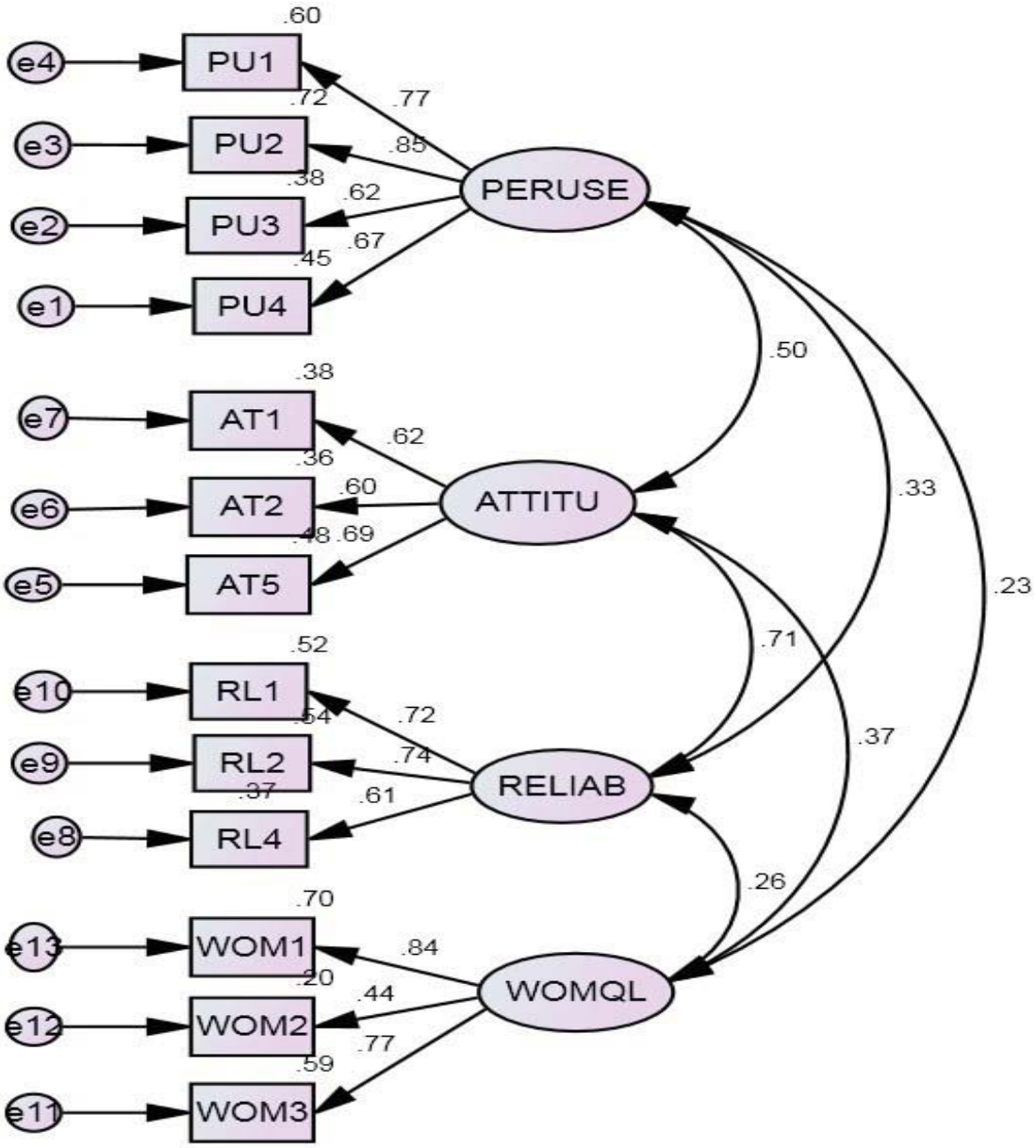
Table 2 Results of EFA

Constructs	Items	Code	Factor Loadings	Cronbach's Alpha
Attitude	I am interested in social media advertisements	AT1	.779	0.675
	I feel comfortable with online advertisements	AT2	.784	
	My attitude toward online advertisement is positive	AT3	.678	
Perceived Usefulness	Social Networking Sites improves my shopping productivity	PU1	.789	0.712
	SNS would be useful in buying what I want	PU2	.801	
	SNS will improve my shopping ability	PU3	.684	
	SNS would enhance my effectiveness in shopping	PU4	.839	
Reliability	SNS's can be trusted to safeguard my personal information	RL1	.737	0.693
	My privacy would be guaranteed on Social Networking Sites	RL2	.756	
	Social Networking sites are reliable	RL3	.635	
WOM Quality	The reviews are relevant for me	WOM1	.613	0.670
	The reviews are helpful	WOM2	.728	
	I get the information I need in the reviews	WOM3	.746	

The proposed model was further analyzed using *Structural Equation Modelling* (SEM) to understand inter-relationships between the variables which were retained after EFA. SEM is a two-step model-building approach which involves testing the *measurement model* which is further followed by testing of hypothesized linkages through *structural model* (Anderson & Gerbing, 1988; Hair, Clark, & Shapiro, 2010).

3.2 Confirmatory Factor Analysis (CFA)

In EFA, there are certain limitations, e.g. items loading on more than one factor, although correlate statistically but they cannot be explained theoretically (Ahire, Golhar, & Waller., 1996). CFA has been recommended by the researchers to overcome the kerbs which EFA suffers from (Lee, 2008; Adil, Akhtar, & Khan, 2013). In the present study CFA was performed using AMOS 20 on the following factors, i.e., Perceived Usefulness (PU), Attitude (AT), Reliability (RL) and Word of Mouth Quality (WOM). Items loading on the respective factors were specified and the measurement model (Figure 2) was then tested for model fit.



(PERUSE = Perceived Usefulness; ATTITU= Attitude; RELIAB= Reliability; WOMQL= Word of Mouth Quality)

Figure 2 Measurement Model

The items were found to be loaded in their corresponding variables and the loadings of all the items were found to be within the acceptable range (Figure 2 and Table 3). The standardized regression weights for all the items emerged to be above the minimum criterion of 0.40 (Ford, MacCallum, & Tait, 1986; Ryu, Han & Jang, 2010). The Chi-square value was found to be 135.657 with 59 degrees of freedom ($p<0.05$).

Table 3 Standardized Regression Weights (CFA)

	Estimate
PU4 ← PERUSE	.671
PU3 ← PERUSE	.618
PU2 ← PERUSE	.846
PU1 ← PERUSE	.772
AT5 ← ATTITU	.694
AT2 ← ATTITU	.604
AT1 ← ATTITU	.618
RL4 ← RELIAB	.606
RL2 ← RELIAB	.736
RL1 ← RELIAB	.721
WOM3 ← WOMQL	.765
WOM2 ← WOMQL	.444
WOM1 ← WOMQL	.837

The value of GFI and AGFI was found to be less than that of the desired value of 0.9 but values of CFI (.916), CMIN/DF (2.299) and RMSEA (.063) were found to be within acceptable range thus the overall model was found to be satisfactory. The overall summary of the key fit statistics for the measurement model is demonstrated in Table 4.

Table 4 Model Fit Indices (CFA)

Fit Index	Recommended Values*	Observed Values
CMIN/DF	< 3.0	2.299
GFI	0.90	.879
AGFI	0.90	.813
CFI	0.90	.916
RMSEA	< 0.070	.063

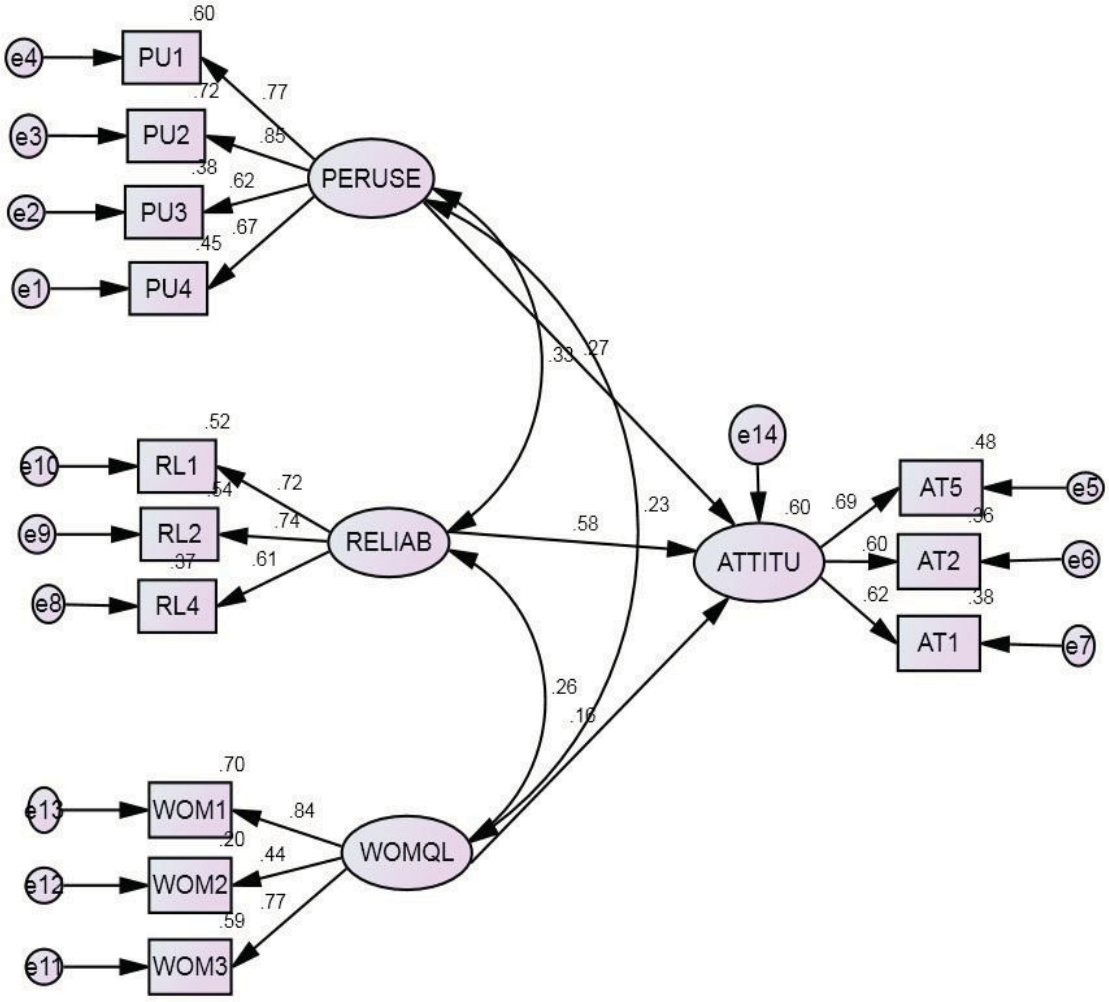
Once the model fit was found to be acceptable, each of the constructs were evaluated for composite reliability (CR), average variance extracted (AVE) and the correlation. The CR of all the constructs were found to be of desirable level ranging from 0.734 to 0.874 (Fornell & Larcker, 1981; Hair, Clark, & Shapiro, 2010; Malhotra & Dash, 2011), indicating adequate reliability of the factors.

Table 5 Validity and Correlation of the Constructs

	AT	PU	REl	WOMQ	AVE	CR
AT	***1***				0.517	0.796

PU	***.628*** **	***1***			0.536	0.820
REL	***.512*** **	***.563***	***1***		0.659	0.874
WOMQ	***.594*** **	***.609*** **	***.576***	***1***	0.494	0.734

The AVE of attitude, perceived usefulness and reliability was well above the minimum value of 0.5, whereas for word of mouth quality the value was found to be 0.494. The values of the AVE indicated adequate convergent validity of the constructs (Fornell & Larcker 1981; O'Leary-Kelly & Vokurka, 1998; Hair, Clark, & Shapiro, 2010; Ryu, Han & Jang, 2010; Khan & Adil, 2013). The CR and AVE values (Table 5) of the constructs confirms the validity and reliability of the scale.



(PERUSE = Perceived Usefulness; ATTITU= Attitude; RELIAB= Reliability; WOMQL= Word of Mouth Quality)

Figure 3 SEM on Proposed Model

3.3 Structural model

The researchers further proceeded with analysis of the proposed model and the hypotheses. The structural diagram of the proposed model is presented in Figure 3. Overall the model fit indices and the goodness of-fit-measures were found to be within the acceptable levels. The value of the GFI (0.891) and AGFI (0.854) were a bit less than that of the desired value of 0.90. The values of CFI (0.908), RMSEA (0.061) and CMIN/DF (2.253) were found to be within the acceptable range. The overall summary of the key fit indices is mentioned in Table 6. Thus, the measurement model was found to be satisfactory.

Table 6 Model Fit Indices for Proposed Model (SEM)

Fit Index	Recommended Values*	Observed Values
CMIN/DF	< 3.0	2.253
GFI	0.90	.891
AGFI	0.90	.854
CFI	0.90	.908
RMSEA	< 0.070	.061

Table 7 and Figure 3 demonstrates the results of the structural model. As shown in Table 7 and Figure 3 the relationship between Attitude and the other three factors, Perceived Usefulness, Reliability and Word of Mouth Quality was found to be significant. Out of the three hypothesized relationships, two of them were found to be significant at significance level 0.01 ($p<0.01$), supporting H1 (Perceived Usefulness → Attitude) and H2 (Reliability → Attitude). However, the relationship of the WOMQL was found to be significant at 0.05 level of significance supporting H3 (WOMQL → Attitude). Although the findings indicate that all the three factors have a positive relationship with the dependent variable (Attitude), but the effect of reliability (.585) on attitude was found to be greater than that of the other two variables.

Table 7 Standardized Regression Weights (SEM)

			Estimates	p- value
ATTITU	←	WOMQL	.146	.039
ATTITU	←	RELIAB	.585	.000
ATTITU	←	PERUSE	.263	.000

4. Discussion & conclusion

4.1 Summary of the study

The purpose of the present study was to examine the determinants of Attitude of the Internet users toward ads over social media. The researchers in the present study have adapted an 18-item scale comprising four variables for use in the Indian settings to explore various dimensions which affects the Attitude. The EFA yielded three constructs (Perceived Usefulness, Reliability and Word of Mouth Quality) which plays a significant role in the formation of positive attitude towards the ads over social media. The results of the study produced a petite scale measuring attitude of the Internet users towards ads over social media.

The EFA and SEM analysis discovered that the proposed model of the present study very finely predicts the attitude of the Internet users towards forming a positive attitude towards ads on the social networks. The convergent and discriminant validity of all the factors of the model were within acceptable range confirming the reliability of the scale. Since all the three proposed hypothesis were supported, it can be surmised that the development of attitude of the Internet users towards advertisements over social networks is determined by PU, RELI and WOMQL. Although all the three factors have a positive impact on the attitude, the results of the present study demonstrate that reliability has the strongest effect on attitude confirming the results of earlier studies (Cheung, Lee, & Thadani, 2009; Safko & Brake, 2009; Akar & Topçu, 2011), followed by Perceived Usefulness and Word of Mouth Quality. Earlier studies have also supported the vitality of PU in the attitude formation (Islam, 2012; Chen & Tseng, 2012; Chow et al., 2012). If the Internet users gets good words about the quality or service provided, they believe the capability of the source (Pai & Huang, 2011). Since all the three variables play an important role in the attitude formation, the marketers aiming to market through social media should pay careful attention to these variables.

4.2 Implications

Social media has become a prime source for the marketers to institute the presence of their brand, this even helps in cutting down the advertising costs (Bolotaeva & Cata, 2010). Social media marketing has been included by almost all the marketers world-wide. To promote products and services the marketers advertise their products and services over SNSs and even have brand pages over the SNSs (Chu, 2011; Tsai & Men, 2013). The presence over the social networking sites can prove advantageous for the marketers as the SNSs users are livelier and their chance of purchasing products online is more. While advertising their products and services over the SNSs, the marketers need to keep in mind the three variables *perceived usefulness*, *reliability* and the *word of mouth quality* which plays an important role in forming a positive attitude of the users towards the advertisement campaign.

5. Limitations and directions for future research

The study may have suffered from certain limitations. The researchers have adopted a convenience sampling and thus the findings of the study may have the limitation of generalizability. There are differences in the sample size of the two group which might have hampered the accuracy of tests applied (Byrne *et al.*, 2007). The generalizability of the results of the research may also have been limited because of the

geographic extent of the study. However, the above mentioned limitations of the study are likely indicators for various other new areas of research. The results of the study can be further validated on a large sample. Future researchers can check the differences in the different groups used for the research.

Various other occupational and educational groups need to be worked upon in the context of advertising over SNSs which might broaden the scope of the findings of the study.

References

Adil, M., Akhtar, A., & Khan, M. N. (2013). Refinement of internet banking service quality scale: a confirmatory factor analysis approach. *International Journal of Services and Operations Management*, *14*(3), 336-354.

Ahire, S. L., Golhar, D. Y., & Waller, M. A. (1996). Development and validation of TQM implementation constructs. *Decision sciences*, *27*(1), 23-56.

Ahmad, A., & Khan, M. N. (2015). Mapping online buyer behavior: A critical review of empirical studies. *Pacific Business Review International*, *8*(2), 37-48.

Ahmad, A., Rahman, O., & Khan, M. N. (2016). Consumer's Perception of Website Service Quality: An Empirical Study. *Journal of Internet Commerce*, 15(2), 125-141.

Ahmad, A., Rahman, O., & Khan, M. N. (2017). Exploring the role of website quality and hedonism in the formation of e-satisfaction and e-loyalty: Evidence from internet users in India. *Journal of Research in Interactive Marketing*, *11*(3), 246-267.

Ahmad, A., & Khan, M. N. (2017). Developing a website service quality scale: a confirmatory factor analytic approach. *Journal of Internet Commerce*, *16*(1), 104-126.

Akar, E., & Topçu, B. (2011). An examination of the factors influencing consumers' attitudes toward social media marketing. *Journal of Internet Commerce*, 10(1), 35-67.

Albanesius, C. (2010). Social networking more popular than email, report says. *PCMag.com. http://www.pcmag.com/article2/0*, *2817*(2367267), 00.

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological bulletin*, 103(3), 411.

Avery, E., Lariscy, R., Amador, E., Ickowitz, T., Primm, C., & Taylor, A. (2010). Diffusion of social media among public relations practitioners in health departments across various community population sizes. *Journal of Public Relations Research*, 22(3), 336-358.

Awad, N. F., & Ragowsky, A. (2008). Establishing trust in electronic commerce through online word of mouth: An examination across genders. *Journal of Management Information Systems*, 24(4), 101-121.

Barefoot, D., & Szabo, J. (2010). *Friends with benefits: A social media marketing handbook*. No Starch Press.

Barnes, N. G., & Mattson, E. (2008). Social media in the Inc. 500: The first longitudinal study. *Journal of New Communications Research*, 3(1), 74-78.

Bausch, S., & Han, L. (2006). Social Networking sites grow 47 percent, year over year, reaching 45 percent of Web users, according to Nielsen/Netratings.

Bennett, S., Maton, K., & Kervin, L. (2008). The 'digital natives' debate: A critical review of the evidence. *British Journal of Educational Technology*, 39(5), 775-786.

Bernoff, J., Pflaum, C. N., & Bowen, E. (2008). *The growth of social technology adoption*. Forrester Research, Cambridge, MA, October, 20, 2008.

Blackshaw, P., & Nazzaro, M. (2006). Word of mouth in the age of the web-fortified consumer. *Consumer-Generated Media (CGM)*, 101.

Bolotaeva, V., & Cata, T. (2010). Marketing opportunities with social networks. *Journal of Internet Social Networking and Virtual Communities*, 2010, 1-8.

Bolton, R. N., Parasuraman, A., Hoefnagels, A., Migchels, N., Kabadayi, S., Gruber, T., Loureiro, Y.K. & Solnet, D. (2013). Understanding Generation Y and their use of social media: a review and research agenda. *Journal of Service Management*, 24(3), 245-267.

Büyüköztürk, Ş. (2003) Sosyal bilimler için veri analizi el kitab ı : İstatistik, araştırma deseni, SPSS uygulamlari ve yorum (Geliştirilmi ş 3.Baskı). Ankara: Pegema Yayınları.

Byrne, B. M., Stewart, S. M., Kennard, B. D., & Lee, P. (2007). The beck depression inventory—II: Testing for measurement equivalence and factor mean differences across Hong Kong and American adolescents. *International Journal of Testing*, 7(3), 293–309.

Cha, J. (2009). Shopping on social networking Web sites: Attitudes toward real versus virtual items. *Journal of Interactive Advertising*, 10(1), 77-93.

Chen, H. R., & Tseng, H. F. (2012). Factors that influence acceptance of web-based e-learning systems for the in-service education of junior high school teachers in Taiwan. *Evaluation and Program Planning*, 35(3), 398-406.

Chen, Y., & Xie, J. (2008). Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management Science*, 54(3), 477-491.

Chen, Y., Wang, Q., & Xie, J. (2011). Online social interactions: A natural experiment on word of mouth versus observational learning. *Journal of Marketing Research*, 48(2), 238-254.

Cheung, C. M., Lee, M. K., & Thadani, D. R. (2009). The impact of positive electronic word-of-mouth on consumer online purchasing decision. In *World Summit on Knowledge Society* (pp. 501-510). Springer

Berlin Heidelberg. Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43(3), 345-354.

Childers, T. L., Carr, C. L., Peck, J., & Carson, S. (2001). Hedonic and utilitarian motivations for online retail shopping behavior. *Journal of Retailing*,77(4), 511-535.

Chow, M., Herold, D. K., Choo, T. M., & Chan, K. (2012). Extending the technology acceptance model to explore the intention to use Second Life for enhancing healthcare education. *Computers & Education*, 59(4), 1136-1144.

Chu, S. C. (2011). Viral advertising in social media: Participation in Facebook groups and responses among college-aged users. *Journal of Interactive Advertising*, 12(1), 30-43.

Clemons, E. K. (2009). The complex problem of monetizing virtual electronic social networks. *Decision Support Systems*, 48(1), 46-56.

Corbett, P. (2009). Facebook demographics and statistics report: 276% growth in 35–54 year old users. Retrieved June, 18, 2009.

DEI Worldwide. (2008). The ımpact of social media on purchasing behavior. Volume One: Initial findings. http://www.deiworldwide.com/files/DEIStudy-Engaging%20ConsumersOnline-Summary.pdf

Dellarocas, C., Zhang, X. M., & Awad, N. F. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing*, 21(4), 23-45.

Dhar, V., & Chang, E. A. (2009). Does chatter matter? The impact of user-generated content on music sales. *Journal of Interactive Marketing*, 23(4), 300-307.

Dobele, A., Lindgreen, A., Beverland, M., Vanhamme, J., & Van Wijk, R. (2007). Why pass on viral messages? Because they connect emotionally. *Business Horizons*, 50(4), 291-304.

Dornyei, Z., & Taguchi, T. (2010). *Questionnaires in second language research: Construction, administration, and processing*. Routledge.

Ford, J. K., MacCallum, R. C., & Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology*, 39(2), 291-314.

Fornell, C., & Larcker, D. F. (1981). Structural equation models with unobservable variables and measurement error: Algebra and statistics. *Journal of Marketing Research*, 382-388.

Fuscaldo, D. (2011). More consumers turn to social media for health care information. (Accessed 29 October 2013).

Garbarino, E., & Strahilevitz, M. (2004). Gender differences in the perceived risk of buying online and the effects of receiving a site recommendation. *Journal of Business Research*, 57(7), 768-775.

Godes, D., & Mayzlin, D. (2004). Using online conversations to study word-of-mouth communication. *Marketing Science*, 23(4), 545-560.

Goh, K. Y., Heng, C. S., & Lin, Z. (2013). Social media brand community and consumer behavior: Quantifying the relative impact of user-and marketer-generated content. *Information Systems Research*, 24(1), 88-107.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W.C. (1998), Multivariate data analysis.

Hair, N., Clark, M., & Shapiro, M. (2010). Toward a Classification System of Relational Activity in Consumer Electronic Communities: The Moderators' Tale. *Journal of Relationship Marketing*, 9(1), 54-65.

Herington, C., & Weaven, S. (2007). Can banks improve customer relationships with high quality online services? *Managing Service Quality: An International Journal*, *17*(4), 404-427.

Heslop, L. (2014). Understanding India: The future of higher education and opportunities for international cooperation. *British Council*.

IAB Internet Advertising Revenue Report. (2009). http://www.iab.net/media/file/IAB-Ad-RevenueFull-Year-2009.pdf (accessed 21 December 2015).

Internetlivestats.com. (2016). Internet users in the world. http://www.internetlivestats.com/internet-users-by-country (accessed 17 April 2016)

Islam, A. K. M. (2012). The Role of Perceived System Quality as Educators' Motivation to Continue E-learning System Use. *AIS Transactions on Human-Computer Interaction*, 4(1), 25-43.

Kelly, L., Kerr, G., & Drennan, J. (2010). Avoidance of advertising in social networking sites: The teenage perspective. *Journal of Interactive Advertising*, 10(2), 16-27.

Kerlinger, F.N. & Lee, H. B. (2000), *Foundations of Behavioral Research*, Harcourt College Publishers, New York, NY.

Khan, M. N., & Adil, M. (2013). Data Analysis Techniques In Service Quality Literature: Essentials and Advances. *Serbian Journal of Management*, 8(1), 95 – 112.

Khare, A., & Rakesh, S. (2011). Antecedents of online shopping behavior in India: An examination. *Journal of Internet Commerce*, 10(4), 227-244.

Kim, M. J., Chung, N., & Lee, C. K. (2011). The effect of perceived trust on electronic commerce: Shopping online for tourism products and services in South Korea. *Tourism Management*, 32(2), 256-265.

Kline, P. (1994). *An Easy Guide to Factor Analysis*, Routledge, London.

Lee, N. (2008). Developing and validating an instrument to assess performance of public sector organisations: a case study of Malaysian schools. *Measuring Business Excellence*, *12*(3), 56-75.

Liao, C., To, P. L., Liu, C. C., Kuo, P. Y., & Chuang, S. H. (2011). Factors influencing the intended use of web portals. *Online Information Review*, 35(2), 237-254.

Liu, Y. (2006). Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of Marketing*, 70(3), 74-89.

Malhotra, N,K., & Dash. S. (2011). *Marketing Research: An Applied Orientation* (6th Ed.). Noida, USA: Pearson.

Marken, G. A. (2007). Social media... The hunted can become the hunter. *Public Relations Quarterly*, 52(4), 9-12.

Mayfield, A. (2008). *What is social media*. http://www.icrossing.com/uk/ideas/fileadmin/uploads/ebooks/what_is_social_media_icrossing_ebook.pdf

Mayzlin, D. (2006). Promotional chat on the Internet. *Marketing Science*, 25(2), 155-163.

Mesure, S., & Griggs, I. (2007). The Facebook betrayal-users revolt over advertising sell-out. *The Independent on Sunday*, 18.

Metin, M., Yilmaz, G. K., Coskun, K., & Birisci, S. (2012). Developing an attitude scale towards using instructional technologies for pre-service teachers. *TOJET: The Turkish Online Journal of Educational Technology*, 11(1), 36-45.

Miller, R., & Lammas, N. (2010). Social media and its implications for viral marketing. *Asia Pacific Public Relations Journal*, 11(1), 1-9.

Muntinga, D. G., Moorman, M., & Smit, E. G. (2011). Introducing COBRAs: Exploring motivations for brand-related social media use. *International Journal of Advertising*, 30(1), 13-46.

Nabi, R. L., Prestin, A., & So, J. (2013). Facebook friends with (health) benefits? Exploring social network site use and perceptions of social support, stress, and well-being. *Cyberpsychology, Behavior, and Social Networking*, 16(10), 721-727.

Nutley, M. (2007). Customer experience is the marketing tool that counts: As more media become interactive and audiences have more control over advertising, marketers must strive to create brand loyalty. *Marketing Week*, 2007, 19.

O'Leary-Kelly, S. W., & Vokurka, R. J. (1998). The empirical assessment of construct validity. *Journal of Operations Management*, *16*(4), 387-405.

Ontario. (2008). Social media marketing: Introduction to social media marketing. http://www.bruce.on.ca/tools/Social_Media_Marketing.pdf (accessed 8 July 2009).

Pai, F. Y., & Huang, K. I. (2011). Applying the technology acceptance model to the introduction of healthcare information systems. *Technological Forecasting and Social Change*, *78*(4), 650-660.

Porter, L., & Golan, G. J. (2006). From subservient chickens to brawny men: A comparison of viral advertising to television advertising. *Journal of Interactive Advertising*, 6(2), 30-38.

Rapp, A., Beitelspacher, L. S., Grewal, D., & Hughes, D. E. (2013). Understanding social media effects across seller, retailer, and consumer interactions. *Journal of the Academy of Marketing Science*, 41(5), 547-566.

Roberts, R. R., & Kraynak J. (2008). *Walk like a giant, sell like a madman*. Hoboken, NJ: Wiley.

Rodgers, S., & Harris, M. A. (2003). Gender and e-commerce: an exploratory study. *Journal of Advertising Research*, 43(03), 322-329.

Ruane, L., & Wallace, E. (2013). Generation Y females online: insights from brand narratives. *Qualitative Market Research: An International Journal*, 16(3), 315-335.

Ryu, K., Han, H., & Jang, S. (2010). Relationships among hedonic and utilitarian values, satisfaction and behavioral intentions in the fast-casual restaurant industry. *International Journal of Contemporary Hospitality Management*, 22(3), 416-432.

Safko, L., & Brake, D. K. (2009). *The social media bible: tactics, tools, and strategies for business success*. Hoboken, NJ: Wiley.

Sen, S., & Lerman, D. (2007). Why are you telling me this? An examination into negative consumer reviews on the web. *Journal of Interactive Marketing*, 21(4), 76-94.

Senthil, M., Prabhu, N. R. V., & Bhuvaneswari, S. (2013). Customers' perception towards advertising in the online shopping and social networking websites among Internet users in India. *AMET, Internafional Journal of Management*, 2(1), 50-59.

Shabnam, S. (2012). The Indian middle class, the state and development: An enquiry into the broad claims of shifts in neo-liberal India. *Inclusive: A Journal of Kolkata Centre for Contemporary Studies*.

Sharma, R. (2008). Bridging the digital divide in Asia-challenges and solutions. *International Journal of Technology, Knowledge & Society*, 1(3).

Shiau, W. L., & Luo, M. M. (2010). Continuance Intention of Blog Users: The Impact of Perceived Enjoyment and User Involvement. In *PACIS* (p. 85).

Sinha, J. (2010). *Factors affecting online shopping behavior of Indian consumers*. University of South Carolina.

Smith, D., Menon, S., & Sivakumar, K. (2005). Online peer and editorial recommendations, trust, and choice in virtual markets. *Journal of Interactive Marketing*, 19(3), 15-37.

Southgate, D., Westoby, N., & Page, G. (2010). Creative determinants of viral video viewing. *International Journal of Advertising*, 29(3), 349-368.

Statista (2016). Distribution of internet users in India as of September 2013, by age group. https://www.statista.com/statistics/272394/age-distribution-of-internet-users-in-india/ (accessed 20 August 2016)

Thoren, E. M., Metze, B., Bührer, C., & Garten, L. (2013). Online support for parents of preterm infants: a qualitative and content analysis of Facebook 'preemie' groups. *Archives of Disease in Childhood-Fetal and Neonatal Edition*, 98(6), F534-F538.

Tsai, W. H. S., & Men, L. R. (2013). Motivations and antecedents of consumer engagement with brand pages on social networking sites. *Journal of Interactive Advertising*, 13(2), 76-87.

Vogt, C., & Knapman, S. (2008). *The anatomy of social networks*.

Weinberg, T. (2009). *The new community rules: Marketing on the social web*. Sebastopol, CA: O'Reilly.

Wellman, B., & Gulia, M. (1999). Net surfers don't ride alone: Virtual communities as communities. *Networks in the Global Village*, 331-366.

Yeh, J. C., Hsiao, K. L., & Yang, W. N. (2012). A study of purchasing behavior in Taiwan's online auction websites: Effects of uncertainty and gender differences. *Internet Research*, 22(1), 98-115.

Yoo, B., & Donthu, N. (2001). Developing and validating a multidimensional consumer-based brand equity scale. *Journal of Business Research*, 52(1), 1-14.

About the authors

Asad Ahmad*
Guest Teacher
Department of Business Administration
Faculty of Management Studies & Research
Aligarh Muslim University Kishanganj Centre, Kishangnanj -855107 (BR) INDIA
Mobile: +91-8791248131
Email: asad7babar@gmail.com

Mohammed Naved Khan
Associate Professor
Department of Business Administration
Faculty of Management Studies & Research
Aligarh Muslim University, Aligarh-202002 (UP) INDIA
Mobile: +91-9557633713
Email: mohdnavedkhan@gmail.com

*Corresponding author

Asad Ahmad did his MBA (Marketing) and PhD (Thesis Submitted) from Aligarh Muslim University, Aligarh, India. Presently he is a guest teacher in the department of Business Administration, Aligarh Muslim University Kishanganj Centre, Kishanganj, India. He has published research papers in journals like Journal of Internet Commerce, Journal of Research in Interactive Marketing and Young Consumers. His research interests include marketing, e-commerce, consumer behavior, and e-consumer behavior.

Mohammed Naved Khan is an Associate Professor at the Department of Business Administration, Aligarh Muslim University, Aligarh, India. He is a recipient of All-India P.D. Agarwal-TCI Award for Doctoral Research in Management (1997) and has co-authored Encyclopedic Dictionary of Marketing published by Sage Publications, USA. His areas of interest include marketing, consumer behavior and business research.

♦

◆

MIS Review 22(1/2), September (2016)/March (2017), 41-58.

Human Resource Information System (HRIS): Re-engineering the Traditional Human Resource Management for Leveraging Strategic Human Resource Management

Abdul Qadir, Swati Agrawal
HR & OB, Jaipuria Institute of Management, Noida, India

ABSTRACT: *Human Resource Management (HRM) functions today no more reside in the comfortable cubicles and sophisticated cabins. In the fast growing business world, the changing paradigms have pushed HRM functions to revive and deliver on the business front. Today HRM has travelled from attendance sheet to balance sheet of the organization in the pursuit of transforming HRM to Strategic Human Resource Management (SHRM). Many HR processes and functions have been reengineered over the years with the help of one of the most important SHRM tools, Human Resource Information System (HRIS). HRIS has leveraged SHRM in more than one ways for HR practitioners enabling the transformation of traditional HRM to transformational HRM. This paper presents a conceptual framework through an HRIS model underpinning the potentials and opportunities that an HRIS offers for practicing SHRM in the twenty first century for the academia and industry professionals.*

KEYWORDS : *HRM, SHRM, HRIS, Strategic Business Partnership, E-HRM.*

1. Introduction

Human Resource Management (HRM) functions in organizations have seen big changes over the years. Being a critical business unit, they no more live in an isolated world and they are now on the roll. Of late, stakeholders have finally realized that HRM functions if aligned to business as strategic partnership, can yield unimaginable results. Today, in fact just not HR, but every business unit, is on board for supporting the strategic mission(s) of the organization. Ashraf at Hewitt-Aon, advocates, *"HR has been playing more direct roles in accomplishing organizational goals, its strategic partnership with business is critical to any business success".* Taking a cue from this view, it is imperative that all HR functions in 21st century have to be linked to strategic HRM practice. So what makes an HRM function strategic in nature? Drucker et al. (1997) have endorsed in their research that today management professionals in general and HR professionals in particular are facing a dilemma of embracing technology along with the challenge of managing HR. It is not the technology, but the challenge of remaining human and managing once humane without which IT tools will yield no big results in long term business. The rapid changes that saw the emergence of Human Resource Information Systems (HRIS) at General Electric in the 1950s was basically came as strategic need. HRM gamut has changed with the advancement of technology in business. HRM moving to its strategic role needed to restructure itself in ensuring that every HR function is customized to a strategic framework, (Hyde & Shafritz, 1977). Sanctis (1986) prompted that as organizations started growing in manpower, the HR operations became more complex in nature. The data of HR got immensely multiplied. Managing HR data and managing the human resources being complimentary to each other called

for an exhaustive system that could manage the whole professional life cycle of employees in organization. Right from HR acquisition to HR exit, there felt a need to see how these traditional HRM practices can be converted into transformational HR practices. Thinking of such a system, besides many other researchers, Tannenbaum in 1990 first defined an HRIS as a system which can be used to acquire, store, manipulate, analyze, retrieve, and distribute pertinent information about human resources in an organization. Tannenbaum's definition gave important impetus for the need of developing an information system which could change the traditional HRM into strategic HRM. Kavanagh, Gueutal, and Tannenbaum (1990) while extending Tannenbaum's definition of HRIS, have further advocated that HRIS philosophy is beyond the integration of computer hardware and associated HR-related software but it strategically focuses on people, forms, policies, procedures, and data, in addition to. This emphasizes on the idea of merely having a computerized information does not fulfill the criteria of managing HR. To manage HR data and servicing the HR needs, there is a strong need of human factor that could integrate the HR process and policies in the system and then manage and deliver HR operations efficiently. Henson (1996) supporting the need of HRIS for key HR decision making, said, technology has to be embraced for the efficient and effective HR operations.

2. Objective & methodology

This study, being exploratory in nature has used qualitative method while reviewing the evolution of HRIS in the context of SHRM. The broad objectives of this study are as under;

1. To understand the development and journey of HRM to strategic HRM
2. To explore the HRIS role in leveraging strategic HRM and
3. To propose a conceptual model of HRIS showing its strategic HRM alignment.

In this pursuit, the secondary data are collected through research papers on the development of HRIS as strategic tool from national and international sources in addition to books on HRIS. In order to enrich the study with primary data, view(s) of industry experts are also taken into consideration about HRIS in action. The in-depth review of research works in the SHRM & HRIS domain that have been referred to, are carefully analyzed over distinguished heads in literature review and then used to develop a conceptual model of HRIS leveraging SHRM. The researches represent global perspectives on the HRIS role in SHRM, predominantly belong to USA, UK, Germany, Middle East and Asian countries.

3. Literature review

3.1 Strategic human resource management (SHRM) & HRIS

Towards 1990s, Human Resource Management (HRM) saw great technological advancement resulting from fast and ever dynamic business environment. As a result, there emerged the concept of Strategic HRM in business. Tracing back, previously, the HRM has tremendously changed as a profession and scientific discipline. The twentieth century wherein employees were treated as just one of the easiest replaceable factors in the organizational system, got changed to one of the key sources of sustainable competitive advantage (Thite, Kavanagh, & Johnson, 2012).

Becker & Huselid (2006) advocated that the organizational performance gets directly affected by human resources as they are now acknowledged as a strategic asset for competitive advantage under the resources-based view of an organization. Arrival and embracement of new technologies has changed HRM functions way too far adding values to the organization's product or service in the realm of HR department as a strategic business partner. Technology today is an enabler to HRM for providing critical technical support for practicing HRM more strategically. Agrawal (2008) echoing with this view explained in her research that previously, external environmental analysis used to pave path for business strategy, while its implementation used to be dependent on internal stakeholders. Today this practice has changed wherein internal capabilities of organizations have become the main driving force for driving competitive advantage and business strategy. This changing perspective has paved path for merging information system with HR practices.

Despite the technical marriage of IT with HR, experts believed that computers are not the panacea or vital most elements in HRIS, but the information. The focus of an exhaustive HRIS remains intact on information, its validity and reliability and ultimately its utility (Kovach & Cathcart, 1999) for users. Teo, Soon, and Fedric (2001) taking a cue from the finding of Kovach study, endorsed that strategic need of an organization is not for administrative purposes or traditional support but majorly for digging out critical HR information which aligns a function of HR with business. The strategic need of HRIS, highlighted by Kovach et al. is well reflected in the findings of Tansley and Watson (2000) who explained that any HRIS project cannot be successfully completed unless and until all the parties like HR, IT, Vendors and Consultants with the much needed and whole hearted support of top management, do not get to the job in a team form.

Dery et al. (2013) in their case study of the organization MFC, based on actor-network approach have found seemingly a different construct for HRIS which proves HRIS as more of a bane than of boon. They have found that potential of HRIS under organization wide ERP has not really produced the achievement of HR transformation. HRIS did not result in improved HR functionality and a more strategic HR function. A highly customized legacy HRIS when overtaken by an HR modules based enterprise wide ERP system, the HR department grossly lost HR functionality and the HR functions tumbled with declined ability to assume strategic HR functionality. Implementing an HRIS in organization is more of strategic decision hence selection of vendors, managing change and careful design and implementation of HRIS become core responsibilities of an HR department (Mohanty & Tripathy, 2007). In another research Dery, Hall, and Wailes (2006) proposed that HRIS is received and perceived as a change catalyst while managing HR therefore users' acceptance and rejection lies on this context at large.

Gautam and Kasuhik (2011) explained in their research that HRIS plays vital role to support SHRM activities such as: developing HR communication, learning at workplace, career management, Business Process Reengineering (BPR) and decision making. Organizations' priorities are now more focused on business and strategic benefits hence lesser traditional gains are expected from HR functions such as recruitment, training and development, performance appraisal and payroll. Strategic opportunities such as team-oriented job designs, HR metrics, quality HR delivery, and incentivizing team performance with the help of HRIS are treated as focus areas (Mehta & Mahajan, 2012). Chauhan, Sharma, and Tyagi (2011) while reflecting upon on strategic alignment of HRIS with HR functions, found that organization-wide functional integration of HR, strategic objectives being met, shapes the objective of having an HRIS in the organization. HR Metrics are yet another important development in SHRM. Calculation of HR Metrics is leveraged by HRIS which also strategically aligns the organizations' business objectives with HR functions

(Thite, Kavanagh, & Johnson, 2012). Inarguably technology today, has changed work stations and HR functions are no exceptions to it. Technology is a key enabler to HRM function as it shifts transactional emphasis to transformational gains under Strategic HRM (SHRM). Automating and streamlining routine HR transactional activities free-up HR executives to focus more on those areas which add value to organization's end product and secure its future (Hendrickson, 2003). Researchers who have constantly endorsed the critical role of HRIS in managing HR strategically have also tried to define HRIS's basic premise. In the league, Alwis (2010) defined Human Resource Information System (HRIS) as "a computerized system used to collect, record, and store, analyze and retrieve data pertaining to an organization's human resources". Parallelly, he also defined the concept of Human Resource Management System (HRMS) as a tool for recruiting, selecting, developing, appraising, and supporting human resources, efficiently and effectively".

3.2 E-HRM and HRIS

Although HR functions have been performed using technology since the time of World War-II, due to increasing competitive pressure, the process got accelerated in 1980s in the wake of cost reduction, improved productivity and advancements in technology leading. The literature on e-HRM and human resource information system (HRIS) recognizes many key advantages of leveraging technology in the delivery of HR services such as: transaction costs reduction, cycle time; streamlining and reengineering HR process, and enhancing the efficiency and effectiveness of HR processes and functions; improving employee satisfaction by delivering quick HR services and shifting the focus of HR from the transaction processing to strategic HRM (Thite, Kavanagh, & Johnson, 2012). In a credible work done by Ruel, Bondarouk, and Looise in 2004, they concluded that e-HRM primarily improves HR's administrative efficiency for cost reduction and strengthens HR policies and processes. Studying e-HRM practices in five companies they established that e-HRM hardly had any positive effect on employee competences, but rather it resulted more in cost reduction and reduced administrative chaos. In an evidence based approach to find the equation between e-HRM and strategic HRM, Marler and Fisher (2013) found that there is no empirical evidence that showed e-HRM serves any direct purpose for strategic outcomes. However, there were some evidences suggested that strategic HRM predicts e-HRM contextually. They concluded that there is a strong need for more empirical studies on e-HRM and strategic HRM to establish meaningful equations between e-HRM and strategic HRM.

According to Kettley and Reilly (2003), a Human Resource Information System (HRIS) based on computers can be called E-HRM which can integrate HR functions. The concept of E-HRM uses an organization-wide network of HR data, information services, databases which are based on HR transactions. E-HRM therefore can be termed as a system based on web and voice technologies for managing HR functions and delivering HR services. Another definition which Gowan (2001) gave of E-HRM is; "a web-based solution that uses latest web application technology to perform real-time HRM functions". She emphasized that even though the E-HRM framework at times appears complex due it comprehensiveness yet due to its user friendly and feature-pack nature, can customized the HR specific needs. e-HRM concentrates primarily on web based and usually stand-alone applications like; e-recruitment and e-learning using social media, while HRIS focuses on an integrated database system encompassing several business modules under an enterprise resource planning (ERP). e-HRM and HRIS both use technology data distributions center, connecting and conducting HR service delivery (Stone & Lukaszewski, 2009). Parry and Tyson (2011) while examining the real outcomes of e-HRM, found that introduction of e-HRM

has definitely enhanced efficiency, service delivery and standardization of some HR functions but with little transformational impact as there was very marginal evidence for involvement of HR in business decision making using e-HRM.

3.3 HR metrics and HRIS

Measuring efficiency and effectiveness is one of the several difficulties of the HR discipline. In contrast to other management functions, HRM function is more delicate, personalized, context-oriented and therefore cannot be managed easily with predefined techniques" (Thite, 2004a). The sophisticated nature of people issues makes it difficult to measure HR's contribution to the company's profitability. It is also important to note that change takes place relatively slowly in HRM (Gratton et al., 1999) hence overnight breakthrough should not be expected. During economic ups and downs, people's expectations and motivation fluctuate and therefore HR policies may be viewed differently in different situations. Hence, short-term gains and quick results as HR effectiveness is vulnerable to mis-representation of real or potential benefits. For example, at Motorola evaluation of training effectiveness is not done with strict and binding parameters, rather it is treated as integral part of overall management improvement (Pfeffer & Veiga, 1999).

HR Metrics and Analytics can be harmful to HRIS products if they are over emphasized. Boudreau and Ramstad (1998) argued that measurement of the HR processes using metrics as bars of efficiency and effectiveness should involve logical, reliable, consistent and flexible parameters. In order to gauge the effectiveness of HR policies, procedures and processes, pure scientific approaches be avoided as they lack human agency approach in most of the current HRIS systems which is detrimental to the success of HRIS (Hesketh & Fleetwood, 2006).

Thite, Kavanagh, and Johnson (2012) in their book Human Resource Information Systems: Basics, applications & directions have talked about HR Metrics that is nowadays increasingly used for SHRM. They have detailed the various HR metrics that leverage the HR information for key decision making through HR processes. Some of these metrics have been described as; absence ratio, absence cost, employee turnover (monthly and annual), cost per hire, yield ratio employee turnover cost (monthly and annual), return on human capital, and HR expense factor. They have deliberated that these metrics are just not mere calculations or crunching numbers about the efficiency of HR operations but these have significant contributions in SHRM. For example, Measurement of absenteeism in organization is done to analyze absenteeism problem and helps in addressing the issue and facilitate management to develop effective attendance and leave policy; yield ratio as a percentage of applicants from a recruitment source that make it to the next stage of the selection process is done to make recruitment process efficient; measurement of costs involved/incurred with a new hire is done to show any substantial improvements to savings/retention costs which helps in streamlining recruitment function efforts; HR expenses in relation to the total operating expenses of the company measures expenses against the budget to analyze effectiveness of HR process that contribute to savings, if any; measurement of ROTI (return on training investment) on employees is done to analyze causes of positive/negative ROI which provides an opportunity to optimize investment with HR practices such recruitment, motivation, training and development; measurement of ROTI (return on training investment) i.e., total financial gains/benefits an organization realized from a particular training program that tells financially whether benefits outweigh the training cost or not; measurement of

employee turnover costs resulting from separation, vacancy, replacement, and training that helps in analyzing the attrition factors and leverages retention efforts. These metrics while at one hand leverage the SHRM objectives while at the other end they streamline HR processes and functions. Once the transactional efficiency is achieved, the transformational efficiency and effectiveness can be integrated or achieved in the process.

3.4 Application of HRIS in transactional & transformational HRM

HRIS can serve a wide range of uses. From simple spreadsheets to complex calculations at ease, it can provide comprehensive HRIS solutions. In the context of management of people, HRIS serves as a tool for managing and solving array of issues and processes in the gamut of HR.

Technology can be used for different functional use within the HR verticals such as; Human Resource Planning, Hiring individuals, performance management, compensation and benefits, training and development, health and safety of employee, legal requirements and employee relations, retention and work-life balance reflected Enshur, Nielson, and Grant-Vallone (2002) in their study. The study by Huo and Kearns (1992) described HRIS holds enormous functional capabilities for managing recruitment and staffing functions. Matching jobs and prospective candidates can be done with ease by a computerized HRIS.

Examining the transitions that HRIS brings in HR processes, Wiblen, Grant, and Dery (2010) shared a social constructivist perspective to link talent management and technology. They emphasized the talent management initiatives must be duly backed up by engaging employees in the organization integrating social setup's requirement which leverage employee's performance. The learning and development verticals hence can be greatly used by HRIS for efficient management of talent.

Obeidat (2012) in his study substantiated that HR functions like forecasting and planning, HR analysis, and communication and are integral functions to core functions of HR like recruitment, induction, appraisal, knowledge management, and records and compliance. He further explained that integral functions ensure the accuracy in the core functions of HRM primarily when the HR transactions are used for strategic purposes. Kundu and Kadian (2012) have discussed that though functional scope of HRIS applications have increased in organizations however main HRIS applications still in maintaining and analyzing employee record and perform pay roll as popular applications. Tyrrell (1999) elaborated that HR professionals can make good use of HRIS for functions like recruitment and selection for generating superior pool of candidates.

An exhaustive HRIS in organization is capable of performing almost all the HR functions right from entry to exit of an employee. To confirm this, Kumar and Mishra (2014) in their research found that reasonably medium and large scale organizations use HRIS for HR planning and analysis; employee and labor relations; health, safety and security issues; HR development; staffing; compensation and benefits and equal employment opportunities.

An extensive study by Towers Perrin (2009) has identified a big bouquet of HR functions that HRIS can deliver, involved; HR & Succession planning, Work force planning, Work force dynamics analysis, Staffing, Applicant recruitment and tracking, Employee data base development, Performance management, Learning and development, Compensation and benefits management, Payroll, Job evaluation, Salary survey and planning and last but not the least International HRM.

Benfatto (2010) in his doctoral dissertation has shared as survey of Cedar Inc. of 2009 on HR Self-service at Baltimore has shown a comprehensive usage HR functions in HRIS across industries which are presented in Table 1 below;

Table 1 HRIS adoption by Industry in 2009

	World wide Average	Agricult Mining Constr	Financial Services	Health -care	Higher Education	High-tech Manufact.	Consume or/Other Manufact.	Other Services	Public Admin.	Retail/ Wholesale	Transp/ Commun./ Public Utility
Administrative	92%	88%	92%	94%	87%	95%	93%	87%	91%	92%	93%
Service Delivery	46%	41%	60%	43%	38%	45%	29%	55%	41%	49%	55%
Workforce Management	43%	31%	40%	48%	30%	33%	41%	52%	27%	48%	46%
Strategic HCM (Talent management)	42%	29%	43%	44%	41%	42%	42%	45%	33%	46%	44%
Business Intelligence	34%	32%	33%	30%	38%	29%	21%	28%	31%	26%	37%

Source: Cedar (2009).

3.5 Industry experts & users view on HRIS

Interviews and discussions with HRIS users and few industry experts who are in the domain of HRIS revealed the following perspectives in the development in HRIS.

Ansari (2012) in an interview vouched for the role of HRP and shared that IBM's paperless online enrollment plan for all of its employees saved the company 1.2 million per year on printing and mailing costs, the employees enjoy working with the online plan. Employees want web access to logon at home rather than through the company intranet. Company worked out a web-based enrollment system that employees and retirees can access from anywhere. Having the system in place, IBM was able to cut costs and give employees the freedom to discover their benefits on their own time and pace.

Murthi (2013) in his interview shared the possibility in the strategic partnership of HR with business due to the increasing contribution of HRIS. He described how the machine is off-loading humans from the burden of managing files and cabinets. The HR team can now devote much time in critical decision over HR of organisation, which was not possible while quoting the HRIS at La Marche' Hypermarket. He observed that HRIS enabled HR team as an active business partner in the center of HR gamut.

Singhal (2014) at NEC Technologies shared that SAP's HCM as the HRMS has integrated critical HR functions using cloud. This cloud technology has eased up the essential dependence on LAN for accessing HR data and now managing HR and taking HR decisions have become much easier. Despite these positives, there are organizations like; KRIBHCO, Le Marche', BIBCOL, AtoZ Info. Solutions, to name a few, are still struggling with their HRIS plan and proposed softwares. The representatives of these organizations unanimously feel that implementing an HRIS is not a joke. Meticulous planning, involvement of key stakeholders duly supported by top management still remains the key for the success of any HR management tool in organization else even a big player like SAP might also struggle for success as the case of KRIBHCO.

3.6 HRIS models & design

Mayfield, Mayfield, and Lunce (2003) while presenting their HRIS model kept organization's mission in the center and deliberated that "HR Planning and Forecasting, HR Analysis, Knowledge Management, HR Communication, Records and Compliance, HR Development factors should be strategically integrated in a comprehensive HRIS. On a little contrary to Mayfield, Teotia (2012) while investigating the success factors of an HRIS proposed his model wherein he segregated transactional HR with strategic HR and said routine HR operations and functions should not be mixed up with the potential of their strategic use even if they are integral in nature. Strohmeier and Kabst (2012) while establishing the success factors of HRIS said, design of HRIS plays a vital role. Identifying two prominent factors for design they said, if managerial functions of HRIS are designed more based on operational needs, it's more valued by HRM professionals and predict the success of HRIS. It meant that if HR functions are available on the click of mouse for mangers it makes things easier to perform. Secondly, line managers and other users of HRIS would love to use HRIS if it is designed in a way that it is available through web access. They further claimed in their study that success of HRIS depends on whether a particular HR function is actually performed or not. Simply dumping a function in HRIS will lead to inefficiency and unexplored potential of HRIS. Shaikh (2012) in while talking about his three HRIS models vis-à-vis HRIS design model, HRIS hexagonal and HRIS phases model, established that the success of any HRIS system is dependent on the design of it, the functional and operational needs are mapped during that phase during the system development life cycle (SDLC) process.

4. Discussion

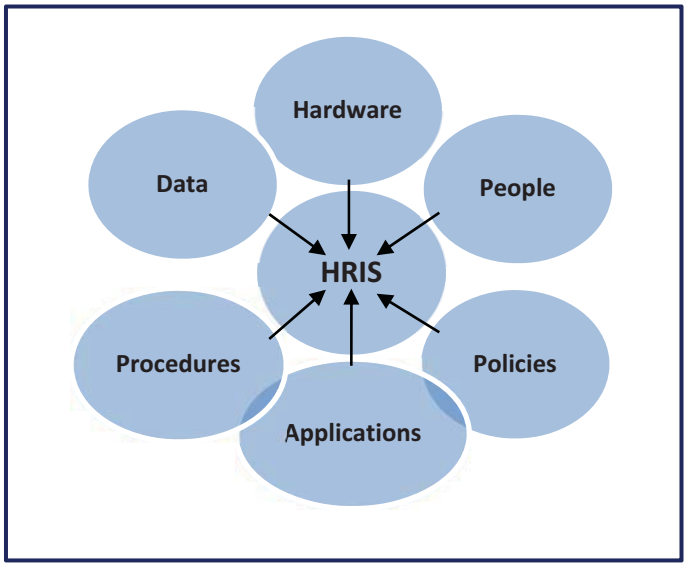
Different researchers have dealt differently with their coverage and objectives of underpinning HRIS as strategic tool under strategic human resource management. To arrive at holistic view and to work-out a conceptual model of HRIS confirming the paper's topic HRIS: leveraging SHRM, the following discussion is being done around the contexts; Strategic Human Resource Management (SHRM) & HRIS, E-HRM & HRIS, HR Metrics & HRIS, Application of HRIS for Traditional & Transformational HRM and HRIS Model & Design;

HRIS since first emerged at GE during 1950's have seen great technological developments specially during 90's. The competitive advantage that it may deliver for sustainable HRM was aptly recognized by HR practitioners as pointed out by Thite, Kavanagh, and Johnson in 2012. Rao (2009) while presenting the historical evolution of HRM has explicitly underlined that how the changing face of traditional HRM has paved path for Strategic HRM. Becker and Huselid (2006) have justly established the resource based view linked to competitive advantage. Agarwal (2008) has evidently highlighted the internal capabilities of HR for bringing great source of competitive advantage rather relying on external factors for the same. The compelling views of Kovach and Cathcart (1999) is true that IT is an enabler for HR operations and therefore should be treated in that context only while using HRIS. This view has been well supported by Teo, Soon, and Fedric (2001). Mohanty and Tripathy (2007), Gautam and Kaushik (2011) have racked in critical HRIS roles of SHRM activities as developing HR communication, learning at workplace, career management, Business Process Reengineering (BPR) and decision making which make the core of SHRM. Chauhan and Tyagai (2011), Mehta and Mahajan (2012) and Thite, Kavanagh, and Johnson (2012) have aptly highlighted the role of HR metrics in making SHRM activities fruitful. However,

Gratton et al. (1999) have cautioned too for not solely relying on numbers arrived via metrics because these might subject to different working conditions. Pfeffer and Veiga (1999), Boudreau and Ramstad (1998), and Hesketh and Fleetwood (2006) have also hinted upon the same precautions.

The evolving forms of HRIS as E-HRM as discussed by Gowan (2001) and by Kettley and Reilly (2003) have reflected the automated HR processes enabling HRIS framework. HR service delivery dimension has also been discussed by Stone and Lukaszewski (2009) in meaningful way.

While unfolding the application of HRIS in transactional and transformational HRM, Huo and Kearns (1992), Wiblen, Grant, and Dery (2010), Obeidat (2012), Kundu and Kadian (2012), Tyrrell (1999), and Kumar and Mishra (2014), all of these researchers have invariably explained the usage of HRIS such as; HR & Succession planning, Work force planning, Work force dynamics analysis, Staffing, Applicant recruitment and tracking, Employee data base development, Performance management, Learning and development, Compensation and benefits management, Payroll, Job evaluation, Salary survey and planning. Benfatto has proven in his doctoral study with the help of Cedar's detailed survey report that has extensively confirmed that HRIS has been adopted by various industries such as; Agriculture, Mining, Construction, Healthcare, Higher Education, Manufacturing, Consumer, Public Admin., Retail, Wholesale, Transport, Telecommunication, Public Utility and Other Services for the HR gamut of Administrative, Service Delivery, Workforce Management, Strategic HCM, Talent Management and Business Intelligence into various percentage of their usage.



Source: This study.

Figure 1 Factors for developing HRIS model

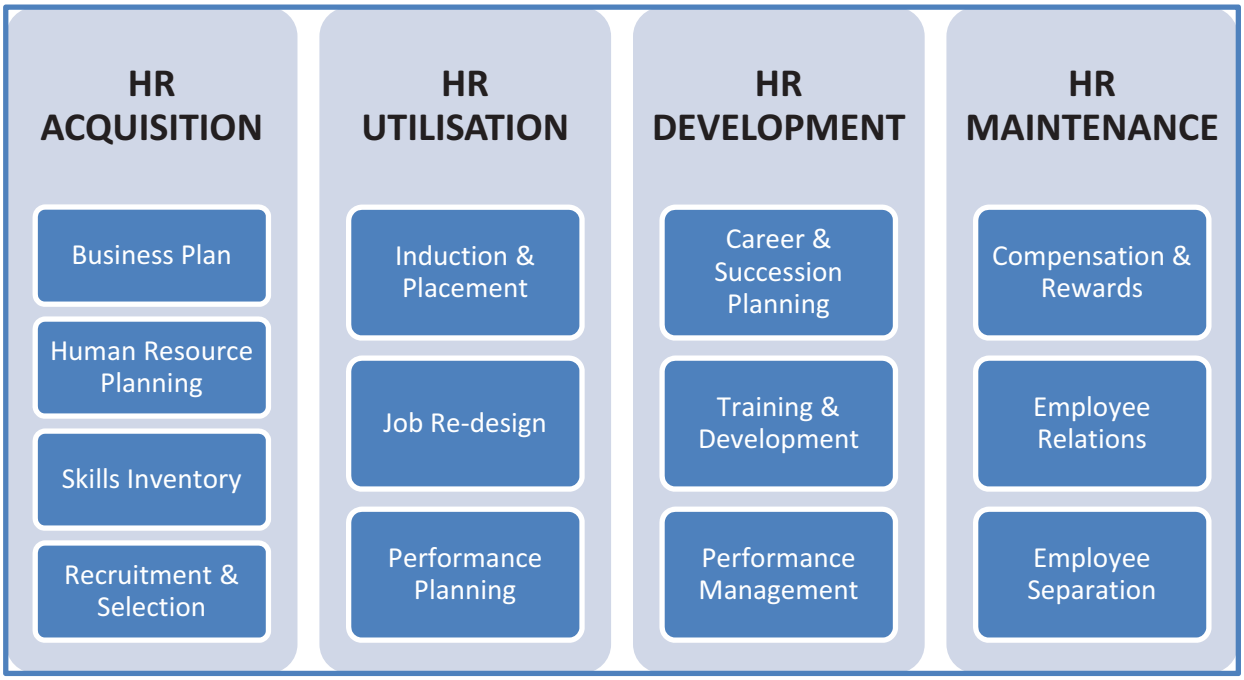
Mayfield, Mayfield, and Lunce (2003), Teotia (2012) and Shaikh (2012) all have given varied factors in the successful development of HRIS models. Their common threads in the development of an HRIS model and its design have been summed up by Hendrickson in 2003. Hendrickson in his research while explaining basic factors for developing an HRIS model, has suggested to create the model around the

important factors as; Hardware, People, Policies, Applications, Procedures, Data and keep the HRIS in the center of it. The same is depicted in Figure 1.

Though these factors exist technically stand-alone yet these are interwoven in nature due to integrated functionality of HRIS. These factors also endorse the HRIS philosophy earlier explained by Kavanagh, Gueutal, and Tannenbaum (1990) that computerized information built around hardware and software is not the key to its success. In addition to these there are factors like; people, forms, policies, procedures, and data that are integral and important factors which ensure the HRIS is properly planned, blue-printed, designed, tested, and finally implemented for the HRIS's transactional and transformational journey for managing human resources.

4.1 Conceptual model of HRIS

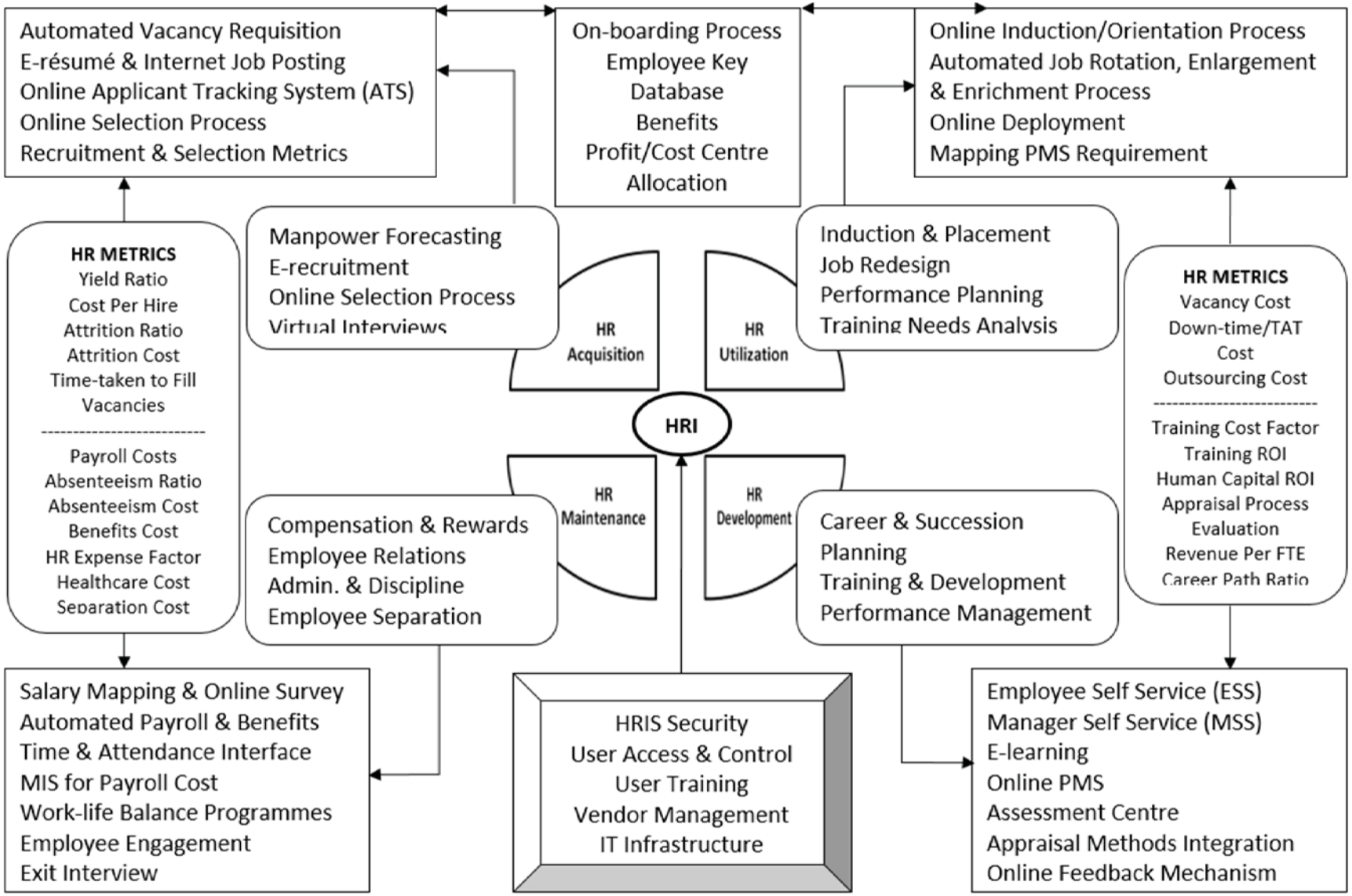
Based on the literature review over thirty-five relevant papers, we have classified the entire gamut of HRM into three heads; Key HR Processes, Transactional and Transformational HR Activities under SHRM. Then each of the key HR processes has been grouped under one specific HR function to further map the transactional and transformational HRM activities under that. The first head i.e., Key HR Processes include the following four subheads; HR Acquisition, HR Utilization, HR Development and HR Maintenance. Further, these subheads are linked to specific HRM & SHRM functions that are presented as the final conceptual HRIS model. The below Figure 2 depicts the key HR Processes which are mapped in Table 2 while Figure 3 represents the conceptual HRIS model.



Source: This study.

Figure 2 Key HR processes

Each of these key HR Processes shown in Figure 3, are now being mapped across traditional HRM functions and transformation SHRM attributes to map in the above model is explained in Table 2 given below. Further, the basic theme of the above model is the marriage of Information Science with HRM, the growing SHRM perspective and the HRIS as a tool and platform to deliver the traditional/transactional and transformational HR functions through a single window HRIS; our conceptual model maps HRM functions in an HRIS around key HR processes. Thite, Kavanagh, and Johnson (2012) have given special emphasis on HR metrics for evaluating the efficiency and effectiveness of HRM functions which paves the path for knowing the success of HRIS adoption through implementing or practicing Strategic HRM in organization. Hence all the above processes are strongly subject to be mapped under strategic angle of achieving alignment with organizational goals. All the functions performed right from employee entry to exit is seen and analyzed through cost benefit equations and use of HR metrics is put to practice to see whether HR processes are yielding positive results or not. Various HR Metrics that are integral part of all the above key HR processes include; Yield Ratio, Cost Per Hire, Attrition Ratio, Attrition Cost, Time-taken to Fill Vacancies, Payroll Costs, Absenteeism Ratio, Absenteeism Cost, Benefits Cost, HR Expense Factor, Healthcare Cost, Separation Cost, Vacancy Cost, Down-time/TAT Cost, Outsourcing Cost, Training Cost Factor, Training ROI, Human Capital ROI, Appraisal Process Evaluation, Revenue Per FTE and Career Path Ratio. These metrics later also form the basis for HR Analytics wherein Descriptive Analysis as HR Metrics pave the path for Predictive and Prescriptive Analysis that are conducted for key decision making related to HR (Fitz-Enz & Matox, 2014).



Source: This study.

Figure 3 Conceptual HRIS

Table 2 Mapping of Key HR Processes with Traditional & Transformational SHRM Functions

Key HR Processes	Traditional HRM Activity	SHRM/HRIS Activity
HR ACQUISITION		
Business Plan	Generalized mapping	Business plan with HR requirements
Human Resource Planning	Paper based manual work	Online & Electronic
Skills Inventory	Paper based manual work	Programmed Method
Recruitment & Selection	Paper resume & paper postings Manual short-listing process	E-resume & internet posting Applicant Tracking System (ATS)
Selection	Costs directed at attraction Manual review of resume Face-to-face interviews Manual Employment Tests	Costs directed at selection E-review of CVs (scanning) Virtual interviews Online Employment Tests
HR UTILISATION		
Induction & Placement	Manual & HR driven process	Automated & line-function driven
Job Re-design	No regular jobs reviews, boring from day one	Job rotation, enlargement and enrichment
Performance Planning	HR & Line manager driven	Superior/Subordinate driven (MBO)
HR DEVELOPMENT		
Career & Succession Planning	HR driven career paths Reactive decision Personal networking	Employees manage their career own Proactive planning with technology Electronic & personal networking
Training & Development	Standardized classroom training HR driven process	Flexible/Virtual/Online training Employee driven process
Performance Management	Supervisor evaluation Face-to-face appraisal Less focus on result based appraisal methods	360 degree evaluation Appraisal software (online/offline) BARS, BOS, Balance Score Card, HR Score Card
HR MAINTENANCE		
Compensation & Rewards	Manual assessment on paper, time consuming Manual/basic payroll automation Internal equity driven HR driven payroll queries	Accurate automated assessment through market analysis Automated payroll or outsourced Internally and externally aligned Online employee self service
Employee Relations	Focus: Employee-Management Union driven process Equal employment opportunity Building & equipment safety Physical fatigue & wellness More task oriented jobs, no focus on work-life	Focus: Employee-Employee Employee driven Intellectual property & data security Ergonomic considerations Mental fatigue & wellness Flexibility through accountability and virtual job performance culture
Employee Separation	A routine process with no focus on employee turnover	Highly focused on attrition and Exit interviews for retention

Source: This study.

5. Conclusion & future scope of work

Summing up the study, the development and journey of HRM to strategic HRM has been promising. The literatures, invariably show that the way human resources were managed three decades ago have seen sea level changes. Expected benefits of introducing machines and softwares for managing HR, have for sure eased-up HR professionals work stations. However, the true predicted paradigm of SHRM is yet to unfold and to be realized. There is a strong need of literature and primary data which should confirm the strategic gains of HRIS in organizations. Case studies either from the company or vendor side will be handy to validate the journey of HRM to SHRM. HRM functions can be immensely leveraged and benefitted by technology, specifically for transactional endeavors and routine HR activities; however, it is transformational tasks and activities that remain at the center of strategic HRM. It is this transformational pursuit that makes HR a strategic business partner to deliver competitive advantage. The social architecture that we come across in literature, endorses the view that technology is mere an enabler for HR and often prone to unpredicted implications. It is ultimately HR who works through technology while understanding the intricacies of data and derived information for managing HR functions at both end, transactional and transformational. Hence, the conceptual model proposed during the discussion, explicitly advises the HR users and professionals to be very careful while working on the integration of IT with HR. One has to be dependent on IT only to an extent which leverages the processes but the wit and knowledge that one can utilize during HR operations, technology unfortunately doesn't cater to. This study is basically a framework for HRIS users who should try to ensure that their operational, transactional needs are sufficiently met through HRIS so that they can reap the benefits of strategic HRIS. The conceptual model proposed in the discussion carries scope of empirical research to see how, this HRIS model can leverage strategic HRM.

References

Agrawal, S. (2008). Paradigm shift in alignment of business strategy and HR strategy. *Fortune Journal of International Management*, 5(2), 49-61.

Ansari, S. (2012, March 27). Role of HR-ERP. (A. Qadir, Interviewer).

Ashraf, M. (2015, Dec. 27). HRIS an Strategic Need. (A. Qadir, Interviewer).

Alwis, D. C. A. (2010). The impact of electronic human resource management on the role of HR managers. *E+M Ekonomika a Management*, 13(4), 47-60.

Becker, B. E., & Huselid, M. A. (2006). Strategic human resource management: Where do we go from here? *Journal of Management*, 32(6), 898–925.

Benfatto, D. M. C. (2010). 'Human Resource Information System and the performance of the Human Resource function', Unpublished Ph.D. dissertation, LUISS Guido Carli, Rome.

Boudreau, J. W., & Ramstad, P. M. (1998). Human resource metrics: Can measures be strategic? *Research in Personnel and Human Resources Management*, Supplement 4, 75-98.

Chauhan, A., Sharma, S. K. & Tyagi, T. (2011). Role of HRIS in improving modern HR operations. *Review of Management*, 1(2), 58-70.

Dery, K., Hall, R., & Wailes, N. (2006). ERPs as 'technologies-in-practice': social construction, materiality and the role of organizational factors. *New Technology, Work and Employment*, 21(3), 229-241.

Dery, K., Hall, R., Wailes, N., & Wiblen, S. (2013). Lost in translation - an actor-network approach to HRIS Implementation. *Journal of Strategic Information Systems*, 22(3), 225-237.

Drucker, P. F., Dyson, E., Handy, C., Saffo, P., & Senge, P. M. (1997). Looking ahead: Implications of the present. *Harvard Business Review*, 75(5), 18–24.

Enshur, E., Nielson, T., & Grant-Vallone, E. (2002). Tales from the hiring line: Effects of the Internet technology on HR processes. *Organizational Dynamics*, 31(3), 224-244.

Fitz-Enz, J., & Matox, J. R. (2014). *Predictive Analytics for Human Resources* (1ed.). New Jersy, Hoboken: John Wiley & Sons.

Gautam, P., & Kaushik, R. (2011). Role of HRIS in strategic HR tasks. *VIVECHAN International Journal of Research*, 2, 117-122

Gowan, M. (2001). *e-HRM: An Internet Guide to Human Resource Management*. New Jersey: Prentice Hall.

Gratton, L., Hope-Hailey, V., Stiles, P., & Truss, C. (1999). Linking individual performance to business strategy: the people process model. *Human Resource Management*, 38(1), 17-32.

Hendrickson, A. R. (2003). Human resource information systems: Backbone technology of contemporary human resources. *Journal of Labor Research*, 24, 381-394.

Henson, R. (1996). HRIMS for dummies. HR Focus, 73(11), 03-05.

Hesketh, A., & Fleetwood, S. (2006). Beyond measuring the HR management-organizational performance link: Applying critical theorist meta-theory. *Organization*, 13(5), 677-699.

Hyde, C. A., & Shafritz, J., M. (1977). HRIS - Tomorrow's system to manage HR. *Public Personnel Management*, 6, 70-77.

Huo, Y. P. & Kearns, J. (1992). Optimizing the job-person match with computerized HRIS. *Personnel Review*, 21(2), 03-18.

Kavanagh, M. J., Gueutal, H. G., Tannenbaum, S. I. (1990). *Human Resource Information Systems: Development and Application*. PWS-Kent Publications Co., Boston, Mass.

Kettley, P., & Reilly, P. (2003). e-HR: An Introduction. Report 398. London: Institute of Employment Studies. ISBN 978-1-85184-326-8.

Kovach, K. A., Cathcart, C. E. Jr. (1999). HRIS providing business with rapid data access information exchange & strategic advantage. *Public Personnel Management*, 28(2), 275-282.

Kumar, S., & Mishra, K. (2014). HRIS - role & significance in todays's global business competitiveness. *Indian Journal of Applied Research*, 4(11), 325-327.

Kundu, S. C., Kadian, R. (2012). Applications of HRIS in HRM in India - A study. *European Journal of Business and Management*, Vol. 4(21), 34-41.

Marler, J. H., & Fisher, S. L. (2013). An evidence based review of e-HRM & SHRM. *Human Resource Management Review*, 23(1), 18-36.

Mayfield, M., Mayfield, J., & Lunce, S. (2003). HRIS - A review & model development. *Advances in Competitiveness Research*, 11(1), 139-152.

Mehta, V., & Mahajan, G. (2012). HRMS - Is the role changing from being administrative to strategic. *International Journal of Computers & Technology*. 3(1), 133-139.

Mohanty, M., & Tripathy, S. K. (2007). Effectiveness of HRIS - a study in NALCO. *Management & Labour Studies*, 32(4), 422-450.

Murthi, M. (2013, April 07). Importance of HRIS. (A. Qadir, Interviewer).

Obeidat, Y. B. (2012). The relationship between HRIS functions & HRM functionalities. *Journal of Management Research*, 4(4), 192-211.

Parry, E., & Tyson S. (2011). Desired goals and actual outcomes of e-HRM. *Human Resource Management Journal*, 21, 335-354.

Pfeffer, J., & Veiga, F. (1999). Putting people first for organizational success. *Academy of Management Executive*, 13(2), 37-48.

Ruel, H., Bondarouk, T., & Looise, J. K. (2004). E-HRM: Innovation or Irritation: An Explorative Empirical Study in Five Large Companies on Web Based HRM. Management Review, 15(3), 364-380.

Sanctis, G. D. (1986). HRIS - A current assessment. *MIS Quarterly*, 10(1), 15-27.

Shaikh, S. (2012). HRIS designing needs for business application. *ZENITH International Journal of Business Economics & Management Research*, 2(1), 256-264.

Singhal, I. (2014, June 10). HRIS leveraging HR functions. (A. Qadir, Interviewer).

Stone, D. L., & Lukaszewski, K.M. (2009). An expanded model of the factors affecting the acceptance & effectiveness of E-HRM systems. *HRM Review*, 19, 134-143.

Strohmeier, S., & Kabst, R. (2012). Evaluating major HRIS system design characteristics - an empirical study. *International Journal of Business Information Systems*, 9(3), 328-342.

Tansley, C., & Watson, T. (2000). Strategic exchange in the development of HRIS. *New Technology, Work & Employment*, 15(2), 108-22.

Teo, T. S. H., Soon, L. G., & Fedric, S. A. (2001). Adoption & impact of HRIS. *Research & Practice in HRM*, 9(1), 101-117.

Teotia, M. K. (2012). Role of HRIS in Performance Evaluation & Decision Making. *International Journal of Multidisciplinary Research*, 2(4), 229-239.

Thite, M. (2004a). *Managing people in the new economy.* New Delhi: Response, Sage.

Thite, M., Kavanagh, M. J., & Johnson, R. D. (2012). Evolution of HRM & HRIS: The role of information technology. In M. J. Kavanagh, M. Thite, & R. D. Johnson (2ed.), *HRISs: Basics, applications & directions* (Pp. 2-34). New Delhi: India, Sage.

Tyrrell, M. W. (1999). Surfing for Success - The Convergence of HRIS & CMC. In Annual Meeting of the Canadian Anthropology Society, Université Laval, Quebec City (Pp. 1-15).

Wiblen, S., Grant, D., Dery, K. (2010). Transitioning to a new HRIS - the reshaping of HR & IT talent. *Journal of Electronic Commerce Research*, 11(4), 251-267

About the authors

Abdul Qadir*

Assistant Professor, HR & OB
Jaipuria Institute of Management, Noida
A-32 A, Sector 62, Institutional Area
Noida 201039 (UP), INDIA
Email: abdul.qadir@jaipuria.ac.in

Swati Agrawal

Professor, HR & OB
Jaipuria Institute of Management, Noida
A-32 A, Sector 62, Institutional Area
Noida 201039 (UP), INDIA
Email: swati.agrawal@jaipuria.ac.in

*Corresponding author

Abdul Qadir is an Assistant Professor of HR & OB at Jaipuria Institute of Management, Noida. He is currently pursuing PhD from Mohan Lal Sukhadia University, Udaipur (Rajasthan), India. He holds an MBA Degree from Jamia Millia Islamia University, New Delhi and is also UGC (NET) certified. He carries an overall 16 years of rich experience out of which more than 8 years in academia and before joining academics, he has been in the industry for over 8 years managing entire gamut of HR; Recruitment, T&D, PMS, Payroll and HR-ERP. He has worked in Manufacturing, Retail, & ITES/BPO industries in various senior-middle level capacities of HR with some of the top companies like Samsung and has also had two overseas assignments in Bahrain & Dubai with world's prominent retail chains. His interests lie in Strategic HRM, HRIS, HR Analytics and Psychometrics. He has published several papers in reputed journals and award winning cases in international conferences on core HR & OB topics.

Swati Agrawal is working as Professor (HR & OB) at Jaipuria Institute of Management, Noida. She has earned her PhD from University of Allahabad and has done MBA with HR specialization. She has 16 years of experience in teaching, research & training and has to her credit more than 30 research papers, cases & articles published in Indian and International Journals in the area of people management practices. Her papers have been cited in International Journals of ADBC category. She is the reviewer of National & International Journals and is in Editorial Board of Indian Journal. Her area of interest includes emerging issues in people management with focus on service sector, strategic human resource management, performance management and leadership. Recently she is undertaking research on working women and career progression.

♦

MIS Review 22(1/2), September (2016)/March (2017), 59-71.

Security modeling tool for information systems: Security Oriented Malicious Activity Diagrams Meta Model Validation

OthmarOthmar Mwambe[1], Isao Echizen[2]
[1]Graduate School of Engineering and Science, Shibaura Institute of Technology, Japan
[2]National Institute of Informatics, Chiyoda, Japan

ABSTRACT: *Various information system security risk management approaches and security modeling languages have been used to address information system security threats. However, the dramatic growth of information system security attacks remains a nightmare. As many other security modeling languages, Mal-Activity Diagrams (MAD) have also been used to model system malicious processes and risk mitigation processes but due to their syntactic and semantic drawbacks, Security Oriented Malicious Activity Diagrams (SOMAD) were introduced in our previous study as an extension of MAD. In this study SOMAD Meta model comprehensiveness and applicability have been validated by using industrial survey. The obtained results show that SOMAD Meta Model is a comprehensive tool enough to address information system security issues at large scope.*

KEYWORDS: *Mal-Activity Diagrams, security modeling languages, Information system security, security risks management, requirements engineering, security risk management approaches, management information systems, security oriented malicious activity diagrams*

1. Introduction

The dramatic growth of ransomware attacks (Scaife et al., 2016) reflects not only the transformation of cyber security attacks towards software applications but also indicates the need for information system security measures adjustment (Mailloux et al., 2016) as they might be inconsistent with modern attack methods (Jasiul et al., 2014). For decades various Information system security management approaches have been used to address information systems security issues, commonly used approaches include Operationally Critical Threat, Asset, and Vulnerability Evaluation methodology(OCTAVE), Security Quality Requirements Engineering (SQUARE), National Institute of Standards and Technology (NIST), CCTA Risk Analysis and Management Methodology (CRAMM), Method for Harmonized Analysis of Risk (MEHARI), Expression of Needs and Identification of Security Objectives methodology (EBIOS) and Information System Security Risk Management Domain Model (ISSRM). In support of such approaches, various security modeling languages have been used to address information system security issues during the designing stage of information systems, Misuse cases Diagrams, Secure Tropos, Mal-Activity Diagrams, and Secure UML are commonly used languages, such tools have played a great role in supporting security requirements definition as well as information system security at large; however, growth of information system security threats remains a challenge (Geiger, 2014).

Security modeling languages are always working with respect to their defined domain meta models and strength of such tools have commonly been evaluated using alignment approaches for instance with ISSRM (Dubois et al., 2010; Sindre,2007; Mwambe, 2013). Mal-Activity Diagrams (MAD) as many other

security modeling tools was previously being aligned with ISSRM and showed some limitations towards ISSRM risk management process coverage (Mwambe, 2013). To address MAD limitations, additional syntaxes were proposed that led into the syntactic and semantic extension of MAD, Security Oriented Malicious Activity Diagrams (SOMAD) (Mwambe & Echizen, 2017). Security Oriented Malicious Activity Diagrams (SOMAD) is a scenario-based approach that relies on its proposed SOMAD Meta Model which is validated in this study. SOMAD Meta Model is designed to handle normal processes, malicious processes as well as risk treatment processes of information system, this property enables it to take advantage over previously used MAD Metal model. SOMAD meta model complies with Information System Security Risk Management Domain Model (ISSRM) and its comprehensiveness towards ISSRM process coverage and applicability have been evaluated and validated by industrial survey the results show that SOMAD is comprehensive enough to address information system security issues at large scope (83.75%).

This paper consists of six sections, related works and background studies have been briefly discussed in the following chapters; however, our contribution starts from section three.

2. Background study and related work

2.1 Information systems Security risk management approaches

These are procedures and written guides that define how security risk management is implemented so as to preserve the objectives of information security.

2.1.1 The Operationally Critical Threat, Asset, and Vulnerability Evaluation (OCTAVE) methodology

Octave is a risk-based strategic assessment and planning technique for security risk management (Alberts et al., 2003). It is a process driven methodology to identify, priorities and manage security risks in two aspects: operational risk and security practices. In OCTAVE approach Security risk management process is completed based on three-phased approach: Build Asset-Based Threat Profiles, Identify Infrastructure Vulnerabilities and Develop Security Strategy and Plans.Octave approach does support small organisations (OCTAVE-s) as well as big organizations (OCTAVE®) and its distinct characterestics are self directed approach and team based (Alberts et al., 2003).

2.1.2 The Security Quality Requirements Engineering (SQUARE) methodology

This approach is focused on requirement engineering process to convey clear understanding of security risk management for information systems. It consists of nine steps and each step identifies participants, inputs, suggested technique and final output. Its process involves interaction of team of requirement engineers and IT project's stake holders. This approach is most effective and accurate when conducted with a team of security expertise with requirement engineers and stake holders of the project (Haley et al., 2008).

2.1.3 National Institute of Standards and Technology(NIST) methodology

This is a risk-based approach for the development of effective risk management process of information systems as it sets basic principles on connection between business drivers and cybersecurity activities (Cybersecurity, 2014). It consists of three parts: Framework Core-identifies cybersecurity activities, impact and guide for support; Framework Implementation Tiers-identifies risk and mitigation process; Framework profile -shows the outcomes with respect to business requirements. It is flexible and it can cover broad security requirements management processes (Cybersecurity, 2014).

2.1.4 CCTA Risk Analysis and Management Methodology (CRAMM)

This is a qualitative security risk management approach developed by UK's Central Computing and Telecommunication Agency(CCTA) (Yazar, 2002). In this approach risk management process is completed through three stages: identification and valuation of assets –defines data, application software and physical assets; Threat and vulnerability assessment-identifies threats and vulnerabilities; and Risk calculation- calculates risk for each asset.

2.1.5 Method for Harmonized Analysis of Risk (MEHARI)

This is risk analysis (RA) and risk management(RM) method based approach developed by French association of information security professionals (Mihailescu, 2012). Risk analysis is completed through five stages: Context establishment, stakes analysis and assets classification, risk identification, risk analysis and risk evaluation. Risk management is complete through four stages: Risk assessment, Risk treatment, risk acceptance and risk communication.

2.1.6 Expression of Needs and Identification of Security Objectives methodology(EBIOS)

This is French central information systems security division(DCSSI) approach which is widely used in public and private sectors. The approach bases on security requirements and objectives of information systems whereby risk analysis process is completed through five stages: context and environmental analysis, security requirements evaluation, risky analysis, identification of risk objectives and determination of security requirements. It is flexible approach as it can easily be adjusted to support other approaches (Hemery et al., 2007).

2.2 Information System Security Risk Management (ISSRM)

ISSRM is a concepts based approach derived from different security related standards (La Rosa & Soffer, 2013). Risk management goes through iterative and continuous process (Figure 1). Risk management process based on asset, risk and risk treatment concepts. Asset-related concepts define things that add value to the organization (information and business assets); Risk-related concepts define components of risk (threat and vulnerability); Risk treatment-related concepts define means of which risk can be mitigated.

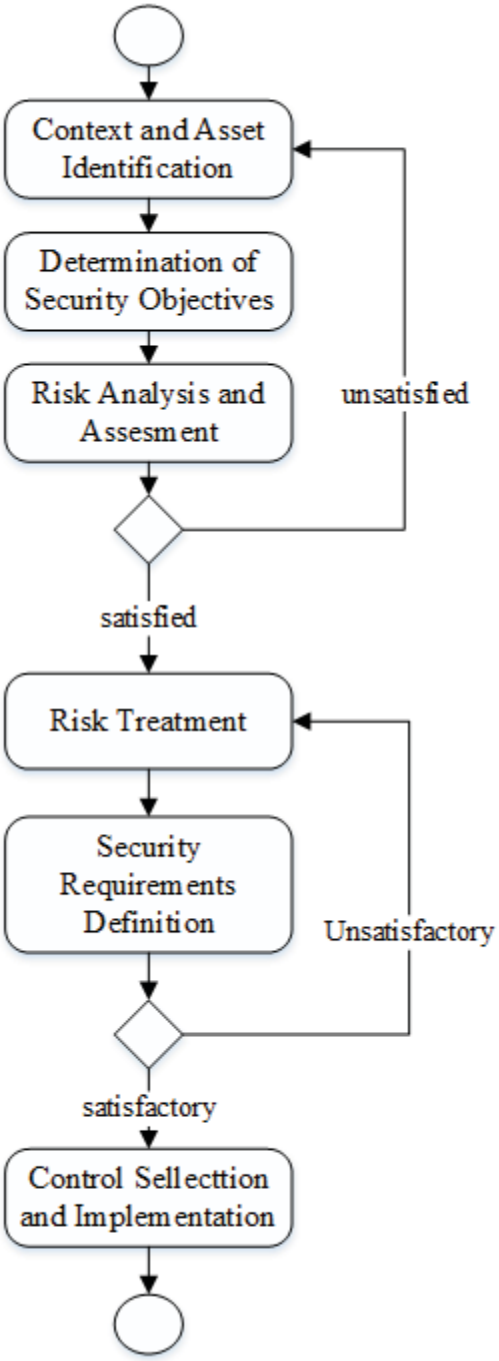


Figure 1 ISSRM Process

Why ISSRM? – Unlike other security risk management approaches, ISSRM is applicable during both information system development and analysis of existing systems (La Rosa & Soffer, 2013) as it integrates risk management process and information system development (Herrmann et al.,2011).This makes ISSRM being suitable for alignment in determination of modeling languages strength (Abbass et al., 2016).

2.3 Related work

Development of security modeling tools with respect to ISSRM process has been an interest of many previous researchers (Mayer, Heymans, & Matulevicius, 2007; Soomro &Ahmed, 2013) due to comprehensiveness of ISSRM domain model. Many previous researchers used ISSRM alignment approach to evaluate their proposed approaches (Chowdhury et al., 2012; Soomro &Ahmed, 2013). Mal-Activity Diagrams and misuse cases both use to model system malicious process, Chowdhury et al. (2012) aligned Mal-activity diagrams with ISSRM to support security requirements definitions, this approach is very useful as it supports software developers during requirement elicitation and designing stage. Mal-Activity is a modified UML Activity (shaded with black color) intended to capture malicious processes, Malicious processes also includes malicious decisions which used by the threat agent (hacker) to harm the system, alignment with ISSRM enables researchers not only to test the strength of the tools but also requirements coverage. Unlike Chowdhury et al. (2012), our approach does not only base on alignment but also semantic extension of Ma-Activity diagrams, SOMAD meta model.

3. Security Oriented Malicious Activity Diagrams (SOMAD)meta model

3.1 SOMAD meta model

Risk management using SOMAD meta model based on three processes: **normal process**, **malicious process** and risk **treatment process**. The Meta Model consists of three main swim lanes: ***Swimlane, Mal-Swimlane*** and ***Control-Swimlane.*** These swim lanes are structured in such a way that One *AnySwimlane* may include many *AnyState*, One *Swimlane* may include many *SwimlaneElements*, one *Mal-Swimlane* may include many *Mal-SwimlaneElements*, one *control-Swimlane* may include many *Control-SwimlaneElements* and all the elements are complete and disjoint (Figure 2).

All processes start with initial state and end with final state, a process may have more than one final state but single initial state. *Swimlane* captures normal information system processes, all normal activities, decisions, security criterion and vulnerabilities of information system are defined by this swim lane; *security criterion* defines security objective of which security requirements (*mitigation activity*) is fulfilled for, while vulnerability defines system weakness that may result into system security breach. Vulnerability is not only the absence of measure; it can also be the existence of an element that makes the system vulnerable to the threats. Malicious processes are captured using *Mal-Swim lane* where all malicious activities and decisions are defined. Risk mitigation processes are captured using *control-Swimlane* where all mitigation activities and decisions to treat are defined. *Control-Swimlane* enables SOMAD metal model to handle risk treatment.

Unlike the meta model in our previous study, current improved meta model has replaced vulnerability with *vulnerable activity*. It has removed security criterion, instead it has defined security criterion as the property of business assets.

3.1.1 Structure flow

Start : SOMAD start with *InitialState* and end with *FinalState*. SOMAD activities are divided into three categories: *Activity*, *Mal-Activity* and *MitigationActivity*. *AnySwimlane* holds all constructs of Security Oriented Malicious Activity Diagrams.

Swimlane includes *SwimlaneElement*, which consists of *VulnerabilityActivity*, *Activity* and *Decision*. *Activity* defines parameterized sequence of behavior. *Decision* defines branching based on either positive or negative conditions. *SecurityCriterion* defines system security objective (security requirement fulfillment) and it is the property of business assets. *VulnerabilityActivity* defines system vulnerability and it identifies activities that can lead to the system security breach.

Mal-Swimlane includes *Mal-SwimlaneElement*, which is composed of *Mal-Activity* and *Mal-Decision*. *Mal-Activity* defines activities performed by threat agent to harm normal process. *Mal-Decision* defines threat agent decision to fulfill malicious goal.

Control-Swimlane includes *Control-SwimlaneElement*, which consists of *MitigationActivity* and *Decision to treat*. *MitigationActivity* defines process improvement to overcome threat. *Decision to treat* defines the decision performed to eliminate the threat.

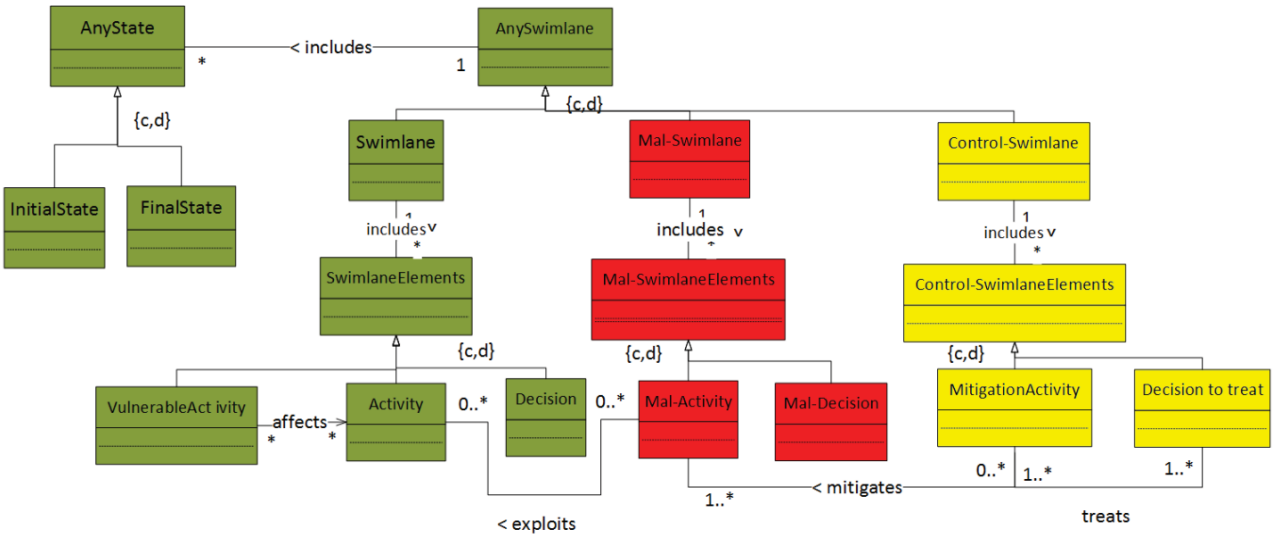


Figure 2 SOMAD meta model

4. Evaluation and validation

SOMAD meta model Comprehensiveness was evaluated based on the analysis of the malicious scenarios applied on test case study “student online information system- SAIS". Evaluation included both alignment approach as well as experimental analysis of the industrial survey. The data was analyzed using IBM SPSS Statistics.

4.1 SOMAD meta model alignment with ISSRM domain model

SOMAD meta model and other security modeling tools were aligned with ISSRM Domain model towards ISSRM risk management process coverage. Out of 13 key security features of ISSRM domain

model, SOMAD meta model could fully address all features of ISSRM domain model (Table1) and took advantage over other modeling tools.

Table 1 Security modeling languages alignment with ISSRM

ISSRM Domain Model		Misuse cases Diagrams Constructs	Secure Tropos Constructs	Secure UML Constructs	Mal-Activity Diagrams Constructs	SOMAD Constructs
Asset-related concepts	Asset	Actor, use cases	Actor, goal, plan, softgoal	Modelelement Class	-	
	Business assets			Class attributes	Activity, decision, control flow	Activity, decision, control flow
	IS assets			Role class, association permission, operation	Swimlane, activity, control flow	Swimlane, activity, control flow
	Security criterion	-	Softgoal,security constraint	-	-	Security constraint
Risk-related concepts	Risk	-	-	-	-	Combination of event and impact
	Threat	Misuser, misuse case	Goal, plan	Role class, association permission	Mal-swimlane,mal-activities, mal-decision	Mal-swimlane,mal-activities, mal-decision
	Threat Agent	misuser	Actor	Role class	Mal-swimlane	Mal-swimlane
	Attack method	Misuse case	Attacks relation, plan	Association permission attributes	Mal-activities, mal-swimlane, mal-decision, control flow	Mal-activities, mal-swimlane, mal-decision, control flow
	Vulnerability	-	Belief	-	-	Vulnerability
	Impact	-	Softgoal and threat combination	-	Mal-activites	Mal-activities
	Event	-	threat	-	-	Combination of threat and vulnerability
Risk treatment-related concept	Risk treatment	-	-	-	-	Combination of control swimlane, mitigation activities,decision to treat and control flow
	Security requirement	Use case	Actor, goal, softgoal, security constraint	Constraint ,constrained elements	Mitigation activity	Mitigation activity
	Control		additional model	additional model	swinlane	Control-swimlane

4.2 SOMAD meta model comprehensiveness validation

Validation based on the findings of the industrial survey that was conducted not only to obtain experts' approval on the comprehensiveness but also to test industrial applicability of the proposed tool. Survey involved 12 experts whereby respondents were provided with detailed description of SOMAD modeling tool and they were supposed to model and give complete analysis of the given scenarios by responding to the questionnaires, response time took around 15-20 minutes. The questionnaires were designed to test the comprehensiveness of SOMAD meta model with respect to ISSRM process coverage.

Scenarios: Hacker launched two attempts to harm student online information system. "Firstly, he flooded student online information system with multiple fake requests; Secondly, hacker used malware to

steal students 'credentials by altering normal flow of online registration process and redirect student to hacker's page".

4.2.1 Demographic information of the respondents and response rate

100% response rate was attained whereby 12 respondents participated in the study, including 7 software developers (66.7%),2(16.7%) system analysts and 2 (16.7%) security experts. 91.7% of respondents were males except one female (8.3%). Regarding education, majority (66.7%) of the respondents were holding masters, many were PhD students, 2 (16.7%) were holding bachelor degree, and 2 PhD (16.7%). All respondents had at least three years working experience in their fields of expertise.

4.2.2 Asset identification

Table 2 shows respondents' rate of response to the questionnaires, "able" shows number respondents who provided correct answers, "not able" false answers and "not sure" are those who left the question blank. 91.7% of respondents successfully identified both business and information system assets (IS).

Table 2 Asset Identification

Result	Frequency	Percentage
Able	11	91.7%
Not able	1	8.3 %
Not sure	0	0%
Total	12	100%

4.2.3 Security objectives determination

75% of the respondents managed to determine security objective defined by the security requirement, two (16.7 %) respondents were not able to determine security objective as they confuse with security requirement and one (8.3%) respondent was not sure (Table 3).

Table 3 Security Objective determination

Result	Frequency	Percentage
Able	9	75 %
Not able	2	16.7 %
Not sure	1	8.3%
Total	12	100%

4.2.4 Risk assessment and analysis

83.3% of respondents were able to identify the vulnerability of information system(SAIS),91.7 % of respondents were able to identify system threat, 83.3% of the respondents were able to identify malicious event, 91.7% managed to identify the attack method used by the threat agent to harm the system and only 75% managed to identify security requirements while 25% failed to identify security requirements (Table 4).

Table 4 Risk assessment and analysis

Result	Frequency			
	Able	**Not Able**	**Not sure**	**Total**
Vulnerability	10 (83.3%)	1 (8.3%)	1 (8.3%)	12 (100%)
Threat	11 (91.7%)	1 (8.3)	0	12 (100%)
Event	10 (83.3%)	1 (8.3)	1 (8.3)	12 (100%)
Attack Method	11 (91.7%)	1 (8.3)	0	12 (100%)
Security requirements	9 (75%)	3 (25%)	0	12 (100%)
Average	**85%**	**11.65 %**	**3.32 %**	

4.2.5 Risk treatment

83.3% of the respondents successfully captured security risk treatment process (Table 5) whereby 83.5% of software developers as well as all (100%) system analysts and all (100%) security experts successfully managed to capture risk treatment process and thus completed ISSRM process (Figure 1).

Table 5 Risk treatment

	Frequency	Percentage
Able	10	83.3 %
Not able	2	16.7 %
Not sure	0	0%
Total	12	100%

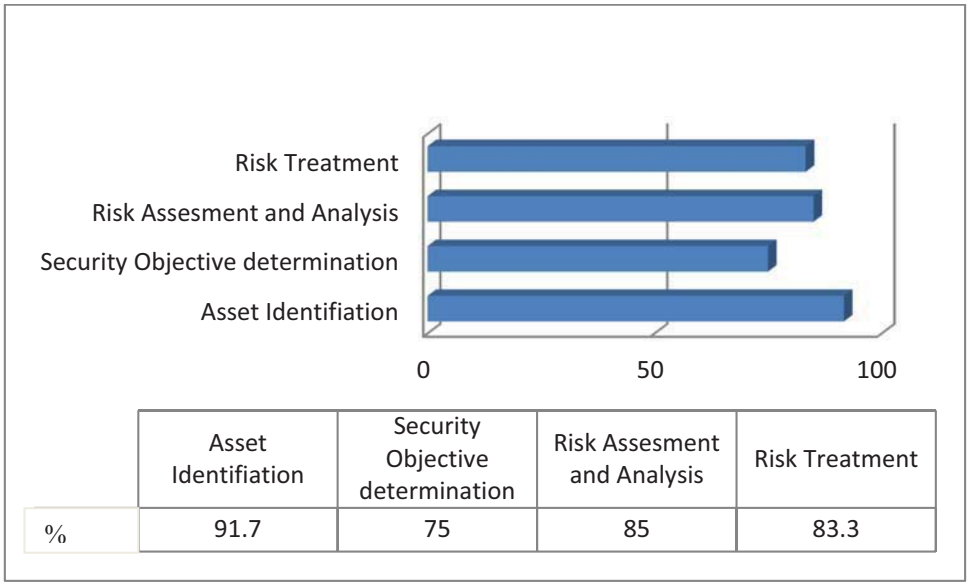


Figure 3 ISSRM Process coverage

4.3 Discussion

Alignment of modeling languages with ISSRM domain model determines the strength of the security modeling tool as it provides clear understanding of the differences and similarities exist in various modeling tools. Regarding the ISSRM risk management process, this wide coverage of ISSRM process gives SOMAD meta model capability to address security issues of information systems at large scope.

Successful asset identification and security objectives determination implies that SOMAD metal model is not only capable of addressing ISSRM assets but also security criterion of information system .Vulnerability, threat and event identification justifies SOMAD meta model is capable of addressing ISSRM risk management process at large scope in turn it can successfully support security analysis of information system. However, 25% of respondents who could not clearly identify security requirements (Table 4) were software developers, due to conflicting requirements elicitation. Thus conflict in requirements elicitation during system design may have significant effect not only on the system but also on system security analysis.

Successful address of risk treatment (Table5) implies that SOMAD metal model is capable of covering ISSRM risk management process at large scope and also reflects the contribution of newly added syntaxes. Huge number of experts successfully managed to capture risk treatment process (Figure 3), this implies that SOMAD can be used not only by software developers but also system analysts and security experts. On average the SOMAD metal model manages to cover 83.75% of ISSRM process (Figure 3), this implies that SOMAD can successfully address information systems security issues at large scope.

5. Managerial Implications

This study suggests that Security Oriented Malicious Activity Diagrams (SOMAD) meta model is industrially applicable and comprehensive enough to address information system security issues at large scope. One of the managerial implications from the results and discussion is the need for the security modeling tools to provide comprehensive framework that can enable software engineers, security experts and system analysts to capture all system security requirements with respect to ISSRM concepts (assets, risk and risk treatment). Fully coverage of ISSRM concepts strengthens the modeling tool as it provides security experts with needful support to address information system security issues.

Risk treatment plays a crucial role in determination of information system security. However, it can only be accomplished successfully if the security requirements and controls are well defined. The introduction of control features and security requirements definition enables software developers and security experts to fully address all system security requirements with respect to ISSRM process.

Non-functional requirements (e.g. availability, vulnerability, confidentiality, integrity etc.) play a crucial role not only in determination of security objectives but also in implementation of organizational security criterion. Having non-functional requirements well defined and emphasized enables software developers, system analysts and security experts to take such requirements into consideration during requirements elicitation, system design and system analysis.

Both information and business assets are very important and characteristics of such assets determine system security requirements. Thus, illustrating features that define such characteristics enable software developers, security analysts and security experts to determine security objectives and implement objective risk treatment process. Such properties also play crucial role in determination of security objectives during system security analysis.

Requirements conflict resolution plays a very crucial role not only during the requirements elicitation but also in the designing of information systems. It is critical success factor in requirements engineering. Based on the obtained result and discussion, having conflicting requirements is inevitable especially when multiple stakeholders are involved. Thus, it is indeed very important to give conflict resolution high consideration as it does have impact on system security analysis.

6. Conclusion

The study has validated comprehensiveness of Security Oriented Mal-Activity Diagrams(SOMAD) meta model toward ISSRM process coverage and also tested applicability of the model. The results show that SOMAD meta model is applicable and comprehensive security tool for management of information systems security, it can be useful tool not only for software developers but also security experts and system analysts. It is easily understood as it provides clear description of the attack and risk mitigation activities. Due to time constraints and nature of the study, the survey could not include large number of respondents as the analysis of the scenarios provided during the industrial survey was time consuming, however majority of the respondents participated in the survey managed to clearly model and analyze risk scenarios and their treatment processes. Thus, SOMAD meta model can address information system security issues at large scope.

References

Abbass, W., Baina, A., & Bellafkih, M. (2016, March). Survey on information system security risk management alignment. In *Information Technology for Organizations Development (IT4OD), 2016 International Conference on* (pp. 1-6). IEEE.

Alberts, C., Dorofee, A., Stevens, J., & Woody, C. (2003). *Introduction to the OCTAVE Approach.* Pittsburgh, PA, Carnegie Mellon University.

Chowdhury, M. J. M., Matulevičius, R., Sindre, G., & Karpati, P. (2012, March). Aligning mal-activity diagrams and security risk management for security requirements definitions. In *International Working Conference on Requirements Engineering: Foundation for Software Quality* (pp. 132-139). Springer, Berlin, Heidelberg.

Cybersecurity, C.I. (2014). Framework for Improving Critical Infrastructure Cybersecurity.

Dubois, É., Heymans, P., Mayer, N., & Matulevičius, R. (2010). A systematic approach to define the domain of information system security risk management. In *Intentional Perspectives on Information Systems Engineering* (pp. 289-306). Springer, Berlin, Heidelberg.

Haley, C., Laney, R., Moffett, J., & Nuseibeh, B. (2008). Security requirements engineering: A framework for representation and analysis. *IEEE Transactions on Software Engineering*, *34*(1), 133-153.

Hemery, H., Akmouche, W., & Tatout, F. (2007). Utilisation de la methodologie EBIOS en securite globale. *REE. Revue de l'électricité et de l'électronique*, (10), 29-125.

Herrmann, A., Morali, A., Etalle, S., & Wieringa, R.J., (2011). RiskREP: risk-based security requirements elicitation and prioritization. In *Perspectives in Business Informatics Research*. Springer Verlag.

Jasiul, B., Śliwa, J., Gleba, K., & Szpyrka, M. (2014, September). Identification of malware activities with rules. In *Computer Science and Information Systems (FedCSIS), 2014 Federated Conference on* (pp. 101-110). IEEE.

La Rosa, M., & Soffer, P. (Eds.). (2013). Business Process Management Workshops: BPM 2012 International Workshops, Tallinn, Estonia, September 3, 2012, Revised Papers (Vol. 132). Springer.

Mailloux, L. O., McEvilley, M. A., Khou, S., & Pecarina, J. M. (2016). Putting the" Systems" in Security Engineering: An Examination of NIST Special Publication 800-160. *IEEE Security & Privacy*, 14(4), 76-80.

Mayer, N., Heymans, P. & Matulevicius, R. (2007). Design of a Modelling Language for Information System Security Risk Management. In *RCIS* (pp. 121-132)

Mihailescu, V. L. (2012). Risk analysis and risk management using MEHARI. *J. Appl. Bus. Inf. Syst.*, *3*, 143.

Mwambe, O. O. (2013). Syntactic and Semantic Extensions of Malicious Activity Diagrams to Support ISSRM. *International Journal of Computer Applications*, *67*(4).

Mwambe, O. O., & Echizen, I. (2017, March). Security Oriented Malicious Activity Diagrams to Support Information Systems Security. In *Advanced Information Networking and Applications Workshops (WAINA), 2017 31st International Conference on* (pp. 74-81). IEEE.

Scaife, N., Carter, H., Traynor, P., & Butler, K.R. (2016, June). Cryptolock (and drop it): stopping ransomware attacks on user data. In *Distributed Computing Systems (ICDCS), 2016 IEEE 36th International Conference on* (pp. 303-312). IEEE.

Sindre, G. (2007, June). Mal-activity diagrams for capturing attacks on business processes. In *International Working Conference on Requirements Engineering: Foundation for Software Quality* (pp. 355-366). Springer, Berlin, Heidelberg.

Soomro, I., & Ahmed, N. (2013). Towards security risk-oriented misuse cases. In *International Conference on Business Process Management* (pp. 689-700). Springer, Berlin, Heidelberg.

Yazar, Z. (2002). A qualitative risk analysis and management tool–CRAMM. *SANS InfoSec Reading Room White Paper*, *11*, 12-32.

About the authors

Othmar Othmar Mwambe*

Graduate School of Engineering and Science, Shibaura Institute of Technology
3 Chome-7-5 Toyosu, Koto, Tokyo 135-8548
Email: mg17007@shibaura-it.ac.jp

Isao Echizen
National Institute of Informatics
Chiyoda, Tokyo, 101-8430, Japan
Email: iechizen@nii.ac.jp

*Corresponding author

Othmar Othmar Mwambe is a graduate student at Shibaura Institute of Technology (SIT). He studied Computer engineering, specialized in Computer Systems and Networks at Kharkiv National University of Radio-electronics (KNURE), Ukraine. He studied software engineering at the University of Tartu (UT), Estonia. Before joining SIT, he was working for Tumaini University, Dar es salaam Institute of Technology (DIT) as instructor and National Institute of Informatics (NII) as a research student. His current research interests include Cyber security, Management Information Systems (MIS), Ubiquitous computing and Brain Computer Interfaces (BCI).

Isao Echizen is a Professor at National Institute of Informatics (NII), Advisor to the Director General (NII). He graduated from the Tokyo Institute of Technology, graduate school of science and engineering. Before joining NII, he worked for Hitachi Ltd as a researcher (Systems Development Laboratory). He is a visiting professor at the University of Freiburg, Halle University, Germany. He is also a Professor at the Graduate University of Advanced Studies (SOKENDAI). His current research interests include media security, media information processing and information Hiding.

♦