Posting Articles for Occupational Stress Reduction in Social Networking Sites: A View of Social Cognitive Theory
Yung-Shen Yen

On Measuring and Increasing the Effectiveness of Banner Advertising
Haren Ghosh, Amit Bhatnagar

Securing DNA Information through Public Key Cryptography
Shiv P. N. Tripathi, Manas Jaiswal, Vrijendra Singh

An Effective Pareto Optimality Based Fusion Technique for Information Retrieval
Krishnan Batri

airiti press

GPN 2007800019
ISSN 1018-1393
DOI: 10.6131/MISR

MIS REVIEW

Volume 19 Number 1 September 2013

MIS REVIEW
An International Journal

Vol.19
No.1 September 2013

MIS Review **Vol. 19, No. 1** Published in September 2013

Publisher: Samuel Chuen-Lung Chen

Published by: National Chengchi University, Department of Management Information Systems & Airiti Press Inc.

Editor-in-Chief: Eldon Y. Li

Executive Editor: Shari S.C. Shang

Assistant Editor: Shu-Hsun Chang

Printed by: Sincere Digital Printing, Co.

Cover Designer: Thomes Chen

Typesetting: Hsiao-Hsuan Wang

Publication Office:

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

Airiti Press Inc.
18F., No. 80, Sec. 1, Chenggong Rd., Yonghe District, New Taipei City 23452, Taiwan

Order Information:

Airiti Press Inc.
18F., No. 80, Sec. 1, Chenggong Rd., Yonghe District, New Taipei City 23452, Taiwan
Tel: +886-2-29266006
Fax: +886-2-22317711
E-mail: press@airiti.com

Price: NT$ 400

e-Journal: http://www.airitilibary.com

ISSN: 1018-1393

GPN: 2007800019

Printed in Taiwan

Editorial Board

Editor's Introduction

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

Yung-Shen Yen in his paper "Posting Articles for Occupational Stress Reduction in Social Networking Sites: A View of Social Cognitive Theory" aims to explore how online users post articles for occupational stress reduction in social networking sites. Drawing on social cognitive theory, this paper examined the effects of subjective norms, personal outcome expectations, and self-efficacy on posting behavior, which in turn reduces occupational stress. A structural equation modeling was used and 262 savvy Facebook users were surveyed. The results revealed that subjective norms, personal outcome expectations, and computer self-efficacy are positively associated with posting behavior, and posting behavior is positively associated with occupational stress reduction. Moreover, the relationship between personal outcome expectations and posting behavior is significant for men, but not for women. In contrast, the relationship between subjective norms and posting behavior is significant for women, yet not for men.

Haren Ghosh and Amit Bhatnagar in their paper "On Measuring and Increasing the Effectiveness of Banner Advertising" argue that banner ad effectiveness can also be determined by measuring the change in perceptions of consumers who have been exposed to a banner ad. They further indicate that the effectiveness of a banner ad can be increased by identifying the issues that are salient to the target consumers and then aligning the message in the banner ad with these issues. A case study is presented where the technique is demonstrated on an advertising campaign launched by the travel department of an Asian country. Consumers who were exposed to the banner ads were shown to be more likely to visit the advertised country

Shiv P. N. Tripathi, Manas Jaiswal and Vrijendra Singh in their paper "Securing DNA Information through Public Key Cryptography" provide robust security to the huge volume of information residing in DNA. In present scenario, security is being managed through symmetric key cryptography only. A new initiative has been taken to increase the robustness of DNA security. In this paper, they are integrating public key cryptography inside traditional DNA security algorithm. The additional security is provided through a new algorithm as proposed, which takes advantage of residue theorem and traditional RSA algorithm. The main security concept is based on complexity in factorization and high versatility of choosing parameters/variables. Basically, DNA is encrypted through symmetric key cryptography and the key used to encrypt the data symmetrically is itself encrypted asymmetrically through proposed modified RSA algorithm. Through examples,

it is further illustrated in this paper that this is not only one of the optimized algorithms to provide a tradeoff between security and computational speed but also adds some sort of defense strategy against various attacks in a layered approach.

Krishnan Batri in his paper "An Effective Pareto Optimality Based Fusion Technique for Information Retrieval" proposes fusion functions to assign relevance scores by considering non dependency among all participating strategies. Relevance score assignment based on the relationship between that specific document and all other documents in the corpus. The existing Comb functions treated as the baseline functions for the proposed functions. Proposed and baseline functions' performance tested among three medium size corpuses. The average precision value of functions indicates that, one of the proposed functions achieves better performance in comparison with the base line functions. The statistical analysis confirms the same.

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. Finally, to our loyal readers around the world, we hope you find the contents of the papers useful to your work or research.

Dr. Eldon Y. Li
Editor-in-Chief and University Chair Professor

Department of Management Information Systems
College of Commerce
National Chengchi University
Taipei, Taiwan
Fall 2013

MIS Review

September 2013 Vol.19 No.1

Contents

Research Articles

MIS Review Vol. 19, No. 1, September (2013), pp. 1-24
DOI: 10.6131/MISR.2013.1901.01

Posting Articles for Occupational Stress Reduction in Social Networking Sites: A View of Social Cognitive Theory

Yung-Shen Yen
Department of Computer Science and Information Management, Providence University, Taiwan

ABSTRACT: *This study aims to explore how online users post articles for occupational stress reduction in social networking sites. Drawing on social cognitive theory, this study examined the effects of subjective norms, personal outcome expectations, and self-efficacy on posting behavior, which in turn reduces occupational stress. A structural equation modelling was used and 262 savvy Facebook users were investigated. The results revealed that subjective norms, personal outcome expectations, and computer self-efficacy are positively associated with posting behavior, and posting behavior is positively associated with occupational stress reduction. Moreover, the relationship between personal outcome expectations and posting behavior is significant for men, not for women. In contrast, the relationship between subjective norms and posting behavior is significant for women, not for men.*

KEYWORDS: *Social Cognitive Theory, Occupational Stress, Social Networking Sites, Facebook.*

1. Introduction

Occupational stress is a growing phenomenon occurring in the workers for the job pressure (King & Gardner, 2006). It can be defined as a taxing when the person appraises the relationship with the work environment (Lazarus & Folkman, 1984). As thus, occupational stress usually is yielded by organizational demands (Cotton & Hart, 2003). Although not all of the outcomes of occupational stress are negative, such as caring nurses work in a demanding work environment (Simmons & Nelson, 2001), but indeed many occupational stresses come from the disorder of the workers to the organization (Nelson & Cooper, 2005). Occupational stress may include work demands and lack of control at work (Karasek, 1979), poor person-environment fit (French, Caplan & Van Harrison, 1982), work-role conflict, role ambiguity or role overload (Kahn, Wolfe, Quinn, Snoek & Rosenthal, 1964), and other factors. For dealing with the stress, people tend to find the stress-reduction ways to relieve the pressure. There are two common approaches to cope with feelings of stress (Burke, 1993; McCarty, Zhao & Garland, 2007). One is positive coping strategies, such as strengthen relations with family members or establish a plan of action to deal with stressful events at work. The other is destructive coping strategies,

such as isolate themselves from friends or family members, increase smoking, or increase consumption of alcohol. However, social support also serves as a buffer against the pressure (Ganster, Fusilier & Mayes, 1986). Talking with friends is another way to cope with the stress. Thus, Internet becomes a suitable outlet for stressors to vent negative emotions. Through the Internet, people can easily post articles regarding the unpleasant work experience to friends for relaxing, especially in social networking sites (SNS), such as Facebook, Twitter, micro-blog, etc.

In literature, social cognitive theory (SCT) has been used for assessing personal behaviors in many different types of research models (Hsu, Ju, Yen & Chang, 2007; Huang & Liaw, 2005; Shih, 2006). According to SCT, personal outcome expectations and self-efficacy are two major factors influencing consumer behavior (Bandura, 1986). Personal outcome expectations refer that people tend to do the work that they believe will result in a better outcome, whereas self-efficacy, also named self-expression in the blog research (Lu & Hsiao, 2009), is concerned with judgments of personal capability. These two factors can be classified as individual's motivations to do the behaviors (Shang, Chen & Shen, 2005). However, for exploring the stressed users to reduce occupational stress through posting articles in SNS, we may change "self-efficacy" as "computer self-efficacy" to represent the capability of using computer for mitigating the stress. On the other hand, subjective norms are classified as a major environmental factor affecting the behavior (Kankanhalli, Tan & Wei, 2005). Subjective norms, as proposed by the theory of reasoned action (TRA), can be defined as perceived social pressure (e.g., peer pressure or superior pressure) to do the behavior (Fishbein & Ajzen, 1975). Lu and Hsiao (2009) found that individuals may publish more information about themselves when they feel that their friends expect them to disclose or publish their information in the blogs. Thus, subjective norms are an important determinant influencing the bloggers to share information. Although we have fully understood the causal relationship for information sharing or purchase intention on the Internet by using the SCT model in information system literature (George, 2004; Lu & Hsiao, 2007; 2009), there are few studies explored stress management to the users through posting behavior in SNS. Indeed, stressed users posting articles in SNS for occupational stress reduction are not only information sharing, but also a psychological behavior (e.g., self-relaxation). That is, excepting for information sharing, SNS may also be a suitable platform for the stressed users to post articles for mitigating the stress. For a stressed user, SNS provides a specific interface to communicate with friends so that he/she can express his/her own emotions or feelings through the platform. In particular, when he/she is depressed, frustrated, or helpless, the user may expect the greetings from the friends or disclose the event by himself on the pages of SNS for venting unpleasant emotions or feelings of stress (Wetzer, Zeelenberg & Pieters, 2007). Thus, it is assumed that users posting behavior in SNS for occupational stress reduction can be influenced by

both environmental factors -- subjective norms, and individual factors -- personal outcome expectations and computer self-efficacy. This study therefore developed an extended framework, incorporating occupational stress reduction into the SCT model, to examine the impacts of subjective norms, personal outcome expectations, and computer self-efficacy on posting behavior, which in turn reduces occupational stress for the users in SNS.

Moreover, stress is subjective and affected by individuals (Gardner & Fletcher, 2009). Research on gender differences in occupational stress has been conducted in the literature (Martocchio & O'Leary, 1989; McDonald & Korabik, 1991). Previous studies suggested that men and women differ in their coping strategies when dealing with stressful situations. For example, McDonald and Korabik found that male workers tend to use the avoidance or withdrawal strategies, while female workers are more likely to talk to others and seek social support than male workers. Burke and Belcourt (1974) argued that women tend to discuss problems with their friends and family more often than men. Thus, it is assumed that the motivations to mitigate occupational stress by posting articles will differ by gender. That is, gender is considered as a moderator in which it may influence the casual relationship of the motivations, posting behavior, and occupational stress reduction.

Therefore, we may ask "Do subjective norms, personal outcome expectations, and self-efficacy influence occupational stress reduction through posting behavior in SNS for the stressed users?" and "Is the causal relationship of the motivations, posting behavior, and occupational stress reduction varied between men and women?" The field has not yet provided direct investigation.

To fill the research gap, we investigated the behaviors of a selected group of stressed users of SNS in Taiwan. It thereby contributes a few significant theoretical results to the field: extending the previous understanding of stress management in the context of SNS; and formulating a united framework to explain the causal relationships for occupational stress reduction across gender. Moreover, it can also help practitioners to extend social functions for users venting feelings of stress in SNS.

2. Literature review and hypotheses development

2.1 Subjective norms affecting posting behavior

Subjective norms refer that an individual believes the person who are important to him/her expect him/her to perform the behavior in question (Fishbein & Ajzen, 1975). Based on the theory of planned behavior (TPB), an extension of TRA, subjective norms are informed by normative beliefs and motivation to comply (Azjen, 1991). For example, a

worker may feel the need to use technology because of the mandate from the organization. Venkatesh and Davis (2000) noted that people may choose to perform a behavior, if they believe the important referents think they should do. Moreover, prior research regarding the adoption of SNS has demonstrated that subjective norms positively affect an individual's IT usage (Hsu & Lu, 2007; Venkatesh & Morris, 2000), and also positively influence the intention to share information (Bock, Zmud, Kim & Lee, 2005; Hsu et al., 2007; Kankanhalli et al., 2005). Bloggers post articles frequently because of peer pressure to blog (Lu & Hsiao, 2009). In a similar way, a worker with occupational stress may be influenced by his/her friends for relaxing the stress. Battacherjee (2000) indicated that subjective norms include external influences such as news reports, the popular press, and mass media. Indeed, SNS provides attractive social features that facilitate users to communicate with others. Thus, when a user finds that SNS members likely post articles for venting negative emotions, he/she may comply with the group norms, and in turn share his/her own experience to the community. Subjective norms, such as peer pressure, can induce users to post articles in SNS for occupational stress reduction. That is, subjective norms will positively influence posting behavior in SNS. This study brings forth the following hypothesis (H1).

H1: Subjective norms are positively associated with posting behavior in SNS.

2.2 Personal outcome expectations affecting posting behavior

Personal outcome expectations refer to the expectations of change in image, status, or rewards (Lu & Hsiao, 2009). Based on SCT, people tend to engage in a behavior if they expect to be rewards (Lu & Hsiao, 2007). Compeau, Higgins and Huff (1999) found that better outcome expectations significantly influence continued usage of information systems. Wasko and Faraj (2005) confirmed that personal outcome expectations significantly affect the intention of using information systems and knowledge sharing. Thus, people likely continue to share information on the Internet if they expect praise or rewards (Lee, Cheung, Lim & Sia, 2006). However, for a stressed worker, the best reward of posting articles in SNS may possibly be greetings, suggestions, or recognition from the friends he/she knew. That is, posting behavior in SNS will be encouraged if a stressed user has a more positive expectation to do that. This study brings forth the second hypothesis (H2):

H2: Personal outcome expectations are positively associated with posting behavior in SNS.

2.3 Computer self-efficacy affecting personal outcome expectations

Computer self-efficacy refers to the confidence in one's ability of using computer on the Internet. Based on SCT, people obtain the confidence in posting articles and

raise self-efficacy when they use computer to share information with others (Lu & Hsiao, 2007). According to this model, self-efficacy positively influences personal outcome expectations, since it is difficult for individuals to separate the consequences of the behavior from their expectations of the outcome (Bandura, 1986). For example, if I believe I will be able to use computer with great skill, I am more likely to expect positive outcomes from my computer use than if I doubt my capabilities (Compeau et al., 1999). Similarly, a user who regularly writes articles in SNS will be an expert with the skill of expressing his/her opinions or feelings to others, and thus have higher outcome expectations than the users who are inability or unfamiliar with the skill. Therefore, self-efficacy, called "computer self-efficacy" in the study, will enhance personal outcome expectations for the users in SNS. This study brings forth the third hypothesis (H3):

H3: Computer self-efficacy is positively associated with personal outcome expectations in SNS.

2.4 Computer self-efficacy affecting posting behavior

Computer self-efficacy also can be recognized as a self-motivational force for the users share the information to others (Jung, Youn & McClung, 2007; Trammell, Tarkowski, Hofmokl & Sapp, 2006). Through posting behavior in SNS, an individual can make himself/herself known to others (Taylor & Altman, 1987). Thus, people tend to use SNS to build relationships or communicate with others through sharing the information or knowledge (Jung et al., 2007). Indeed, computer self-efficacy is a primary motivation in the use of SNS (Lu & Hsiao, 2009; Trammell & Keshelashvili, 2005). It is possible that a stressed user posting articles in SNS may just only express the unsatisfied experience about the work to the public or he/she likes to announce the experience to let others know him/her. Like to tell the parents when a child has been bullied, people tend to tell close friends about the unpleasant experience of the work for releasing the emotion. Thus, it is assumed that computer self-efficacy will positively influence posting behavior for the users in SNS. This study brings forth the fourth hypothesis (H4):

H4: Computer self-efficacy is positively associated with posting behavior in SNS.

2.5 Posting behavior affecting occupational stress reduction

Based on TPB, a person's performance of a certain behavior is determined by his/her intent to perform that behavior (George, 2004). According to the theory, motivation to comply with the views of important others, and attitudes about the services, will influence intent to make the purchases (Azjen, 1991). Thus, posting behavior in SNS can be a trigger for the users to vent the emotions. The behavior is similar to smoking or listening music, which can release the pressure for the users. It is expected that the relationship between posting behavior and occupational stress reduction will be tightly connected if the

users believe that posting behavior can be helpful to mitigate occupational stress. Thus, it is assumed that posting behavior positively affects occupational stress reduction in SNS. This study brings forth the fifth hypothesis (H5):

H5: Posting behavior is positively associated with occupational stress reduction in SNS.

2.6 The moderating effect of gender

In literature, the moderating effect of gender has been considered because there are differences between men and women in terms of communication styles and usage behavior of the internet (Herring & Paolillo, 2006; Stowers, 1995). For example, Gefen and Straub (2000) found that women tend to use electronic communication for building the rapport, but men tend to use it for reporting. Also, Herring (1996) indicated that requesting and providing information are more common for women than for men. Lu and Hsiao (2009) found that personal outcome expectations more strongly determined usage intention for men than for women, but self-efficacy has a more salient effect on usage intention for women than for men. Regarding stress management research, scholars found that women and men have different coping strategies in dealing with stress (McCarty et al., 2007). For example, Davidson and Cooper (1983) argued that women managers dealing with stress are more likely to talk to someone they knew than male managers. Folkman and Lazarus (1988) noted that when individuals encounter stressful situations, men tend to engage in problem-focused coping more often than women. Lim and Teo (1996) found that female IT personnel are more likely to seek social support than male IT personnel in dealing with stress. In contrast, male IT personnel are more likely to engage in an objective and unemotional manner to deal with stress. Thus, it is speculated that women are more social-oriented while men are more task-oriented in dealing with stress. Venting feelings of stress in SNS may result from different motivations across men and women. Therefore, it is expected that gender moderates the causal relationship of the motivations, posting behavior, and occupational stress reduction in SNS. In other words, subjective norms, personal outcome expectations, and self-efficacy may have different impacts on posting behavior, which in turn influences occupational stress reduction between men and women. This study brings forth the sixth hypothesis (H6):

H6: Gender moderates the causal relationship of the motivations, posting behavior, and occupational stress reduction in SNS.

3. Research method

3.1 Framework of the research

Figure 1 depicts the research framework of this study, in terms of the literature established before.

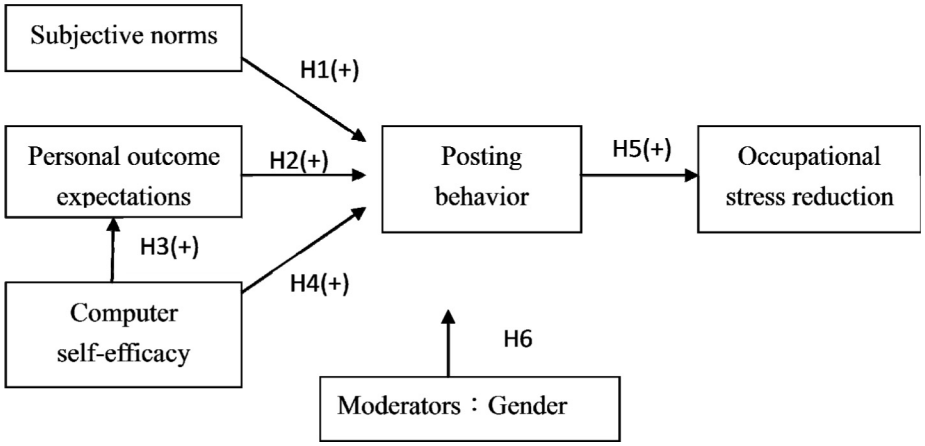

Figure 1 Research Framework of this Study

3.2 The instrument of the study

The design of the instrument is adopted from the previous results in the literature with appropriate modifications for Facebook users (e.g., the terminologies). The scale of subjective norms, personal outcome expectations, and self-efficacy were revised from Lu and Hsiao (2009), which includes two items, three items, and three items, respectively. The scale of posting behavior was revised from Wu (2006), which includes two items. The scale of occupational stress reduction was revised from McCarty et al. (2007), which includes eight items. All of the items were measured on a 7-point Likert-type scale, where possible answers ranged from strongly disagree (1) to strongly agree (7). The instrument of the study is shown in Appendix.

3.3 Subjects

As the sample of this study is from Taiwan, the questionnaire was translated by a language professional to ensure that the wording used in the Chinese and English versions were consistent. This study conducted a convenient sampling. Four assistants invited the respondents who have the experience of posting behavior in the page of Facebook for occupational stress reduction through their social networks in Facebook. To increase the sample return rate, this study offered gifts to the respondents for increasing the participation to the survey.

The questionnaires were collected for one month. A total of 277 respondents were received, of which 15 copies were deleted due to regular or incomplete data. The valid respondents totaled 262. The demography of the respondents is shown in Table 1. Females (53.1%) surpass males (46.9%). The largest age group is 20 ~ 29 years (44.7%), and the largest education group is undergraduate (62.6%). Regarding the frequency of using the internet, most of the respondents use the internet once a day (85.9%).

3.4 Reliability and validity test

This study employed Cronbach's alpha (α) for examining the internal consistency of the constructs (Nunnally, 1978; Robert & Wortzel, 1979). The a in Table 2 indicates the reliability of the measurement constructs: subjective norms are 0.89, personal expectations outcome is 0.90, computer self-efficacy is 0.80, posting behavior is 0.85 and occupational stress reduction is 0.93. These numbers satisfy the general requirements in the field (e.g., Nunnally) suggest a reliability coefficient above 0.7, and Robert and Wortzel want the number to be between 0.70 and 0.98. Therefore, we content that this study carried good reliability.

Confirmation factor analysis (CFA) was performed for scale validity assessment (Anderson & Gerbing, 1988). Convergent validity was measured by average variance extracted (AVE) in each construct. The criterion of AVE should be greater than 0.5 (Fornell & Larcker, 1981). As shown in Table 2, all constructs were satisfied. Thus, this study possessed adequate convergent validity.

Table 1 Demography of the Respondents

Variables	Items	*N*	Per cent (%)
Gender	Male	123	46.9
	Female	139	53.1
Age	20 ~ 29	117	44.7
	30 ~ 39	71	27.1
	40 ~ 49	48	18.3
	> 49	26	9.9
Education	Senior high school	49	18.7
	Undergraduate	164	62.6
	Graduate	49	18.7
Frequency of using the internet	Once a day	225	85.9
	Once a week	31	11.8
	Once a month	3	1.1
	Over one month	3	1.1

Note: valid samples = 262.

Table 2 Model of Research Construct

Construct and Observable Variable	Mean (SD)	SFL	CR	AVE	α
Subjective norms (SN)			0.89	0.81	0.89
SN1	4.79 (1.45)	0.92			
SN2	4.98 (1.43)	0.87			
Personal outcome expectations (POE)			0.90	0.76	0.90
POE1	4.55 (1.51)	0.80			
POE2	4.67 (1.59)	0.89			
POE3	4.39 (1.58)	0.90			
Computer self-efficacy (CSE)			0.78	0.54	0.80
CSE1	4.08 (1.77)	0.74			
CSE2	4.51 (1.58)	0.78			
CSE3	4.11 (1.54)	0.68			
Posting behavior (PB)		0.85	0.87	0.70	0.85
PB1	4.57 (1.56)				
PB2	4.51 (1.49)	0.91			
Occupational stress reduction (OSR)			0.93	0.64	0.93
OSR1	4.58 (1.62)	0.87			
OSR2	4.48 (1.50)	0.91			
OSR3	4.56 (1.52)	0.87			
OSR4	4.43 (1.50)	0.88			
OSR5	4.21 (1.64)	0.78			
OSR6	4.09 (1.58)	0.73			
OSR7	4.59 (1.75)	0.55			
OSR8	3.58 (1.66)	0.70			

Discriminate validity was also tested. The result shows that the AVE square root of each research variable is larger than the related coefficients of the variables, as shown in Table 3. This is a clear case of positive proof (Fornell & Larcker, 1981). Thus, this study had adequate discriminate validity.

3.5 Measurement invariance tests

In order to compare the causal relationship of the motivations, posting behavior, and occupational stress reduction between men and women, this study conducted multiple-group confirmatory factor analysis for testing measurement invariance across gender. The degree of invariance is frequently assessed by the differences in χ^2 between the models (Cheung & Rensvold, 2002). If $\Delta\chi^2$ is not statistically significant, then the invariance

Table 3 Correlation between Constructs

	CSE	SN	POE	PB	OSR
CSE	**0.74**				
SN	0.60	**0.90**			
POE	0.58	0.38	**0.87**		
PB	0.61	0.48	0.49	**0.84**	
OSR	0.62	0.50	0.52	0.72	**0.80**

Note: Diagonal elements in boldface represent the square root of AVE.

exists. Table 4 shows the results of measurement invariance tests. The evidence reveals that factorial invariance (i.e., same factor loadings across gender) and structural invariance (i.e., same factor loadings and factor covariance across gender) both are not significant between men and women ($\Delta\chi^2(14) = 10.097$, $p > 0.05$; $\Delta\chi^2(10) = 15.011$, $p > 0.05$), but error invariance (i.e., same factor loadings, factor covariance, and error variance across gender) are significant ($\Delta\chi^2(19) = 56.762$, $p < 0.05$). This result implies that factorial invariance model and structural invariant model are invariant, but error invariance model is non-variant. However, it is widely accepted that the test of error variance and their covariance represents an overly restrictive test of the data (Byrne, 2010, p. 199). Thus, we believe that measurement invariance does exist in this study.

4. Analysis of empirical results

4.1 Verification of the hypotheses

A structural equation modeling using AMOS 20.0 was conducted to test the postulated hypotheses. Figure 2 presented the estimation results. From the model fitness indexes, $\chi^2(146) = 374.970$, GFI = 0.942, AGFI = 0.915, CFI = 0.974, RMSEA = 0.052, showing the collected data fits the postulated model. The estimated structural coefficients were used to test each hypothesis. Results in Table 5 showed that the model explained 46.7% of the variance in personal outcome expectations, 54.5% of the variance in

Table 4 Results of Measurement Invariance Tests

Model	NPAR	χ^2	DF	$\Delta\chi^2$	ΔDF	*p*	ΔTLI	ΔCFI
Base model	91	1325.093	289					
Factorial invariance	77	1335.190	303	10.097	14	.755	-.012	-.001
Structural invariance	67	1350.201	313	15.011	10	.132	-.006	.001
Error invariance	48	1406.963	332	56.762	19	.000	-.005	.006

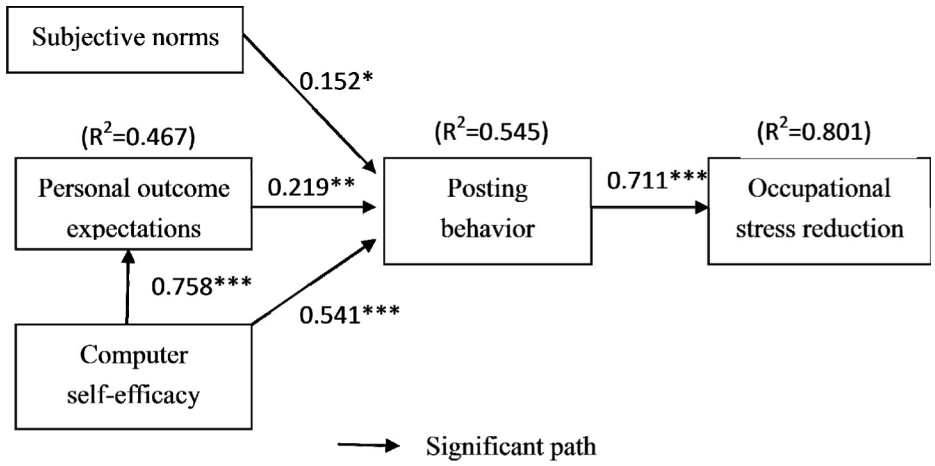

Figure 2 Results of Structural Modelling Analysis

* means significant at the level of 0.05; ** means significant at the level of 0.01; *** means significant at the level of 0.001.

Table 5 Results of Estimated Structural Coefficients

Relationship	Estimate	SE	CR	*p*	Result
H1. SN → PB	0.152	.075	2.035	0.042	Support
H2. POE → PB	0.219	.074	2.946	0.003	Support
H3. CSE → POE	0.758	.085	8.884	***	Support
H4. CSE → PB	0.541	.124	4.348	***	Support
H5. PB → OSR	0.711	.054	8.739	***	Support

Note: Estimate is unstandardized.

*** means significant at the level of 0.001.

posting behavior, and 80.1% of the variance in occupational stress reduction. All paths in the research model were statistically significant at the level of 0.05. Subjective norms, personal outcome expectations, and self-efficacy are positively associated with posting behavior (Estimate = 0.152, SE = 0.075, CR = 2.035, $p = 0.042$; Estimate = 0.219, SE = 0.074, CR = 2.946, $p = 0.003$; Estimate = 0.541, SE = 0.124, CR = 4.348, $p < 0.001$). Moreover, self-efficacy is positively associated with personal outcome expectations (Estimate = 0.758, SE = 0.085, CR = 8.884, $p < 0.001$), and posting behavior is positively associated with occupational stress reduction (Estimate = 0.711, SE = 0.054, CR = 8.739, $p < 0.001$). Thus, H1, H2, H3, H4, and H5 are supported.

Multiple group structural equation modeling was conducted to test the differences of the causal relationships across two groups. Significant differences in two groups were determined by using a χ^2 difference test (e.g., Yang & Lee, 2010). Thus, Table 6 showed

Table 6 The Moderating Effect of Gender

Path	χ^2	DF	*p*
Full constrained model	12.832	5	0.025*
Path 1: SN → PB	4.598	1	0.032*
Path 2: POE → PB	6.508	1	0.011*
Path 3: CSE → POE	0.615	1	0.433
Path 4: CSE → PB	1.258	1	0.262
Path 5: PB → OSR	0.126	1	0.722

* means significant at the level of 0.05.

that full constrained model with limited the same path coefficients between men and women is significant ($\Delta\chi^2/5 = 12.832$, $p = 0.025$). Moreover, the path between subjective norms and posting behavior and the path between personal outcome expectations and posting behavior both are significant ($\Delta\chi^2/1 = 4.598$, $p = 0.032$; $\Delta\chi^2/1 = 6.508$, $p = 0.011$), but other paths are insignificant across men and women. For more details, Figure 3 and Figure 4 showed the results of estimated structural modeling analysis for male and female respectively. The path between subjective norms and posting behavior is stronger for women (Estimate = 0.225, SE = 0.094, CR = 2.393, $p = 0.017$) than for men (Estimate = 0.030, SE = 0.112, CR = 0.269, $p = 0.788$), as shown in Table 7. In contrast, the path between personal outcome expectations and posting behavior is stronger for men (Estimate = 0.440, SE = 0.104, CR = 4.222, $p < 0.001$) than for women (Estimate = 0.065, SE = 0.102, CR = 0.637, $p = 0.524$). Thus, H6 is supported.

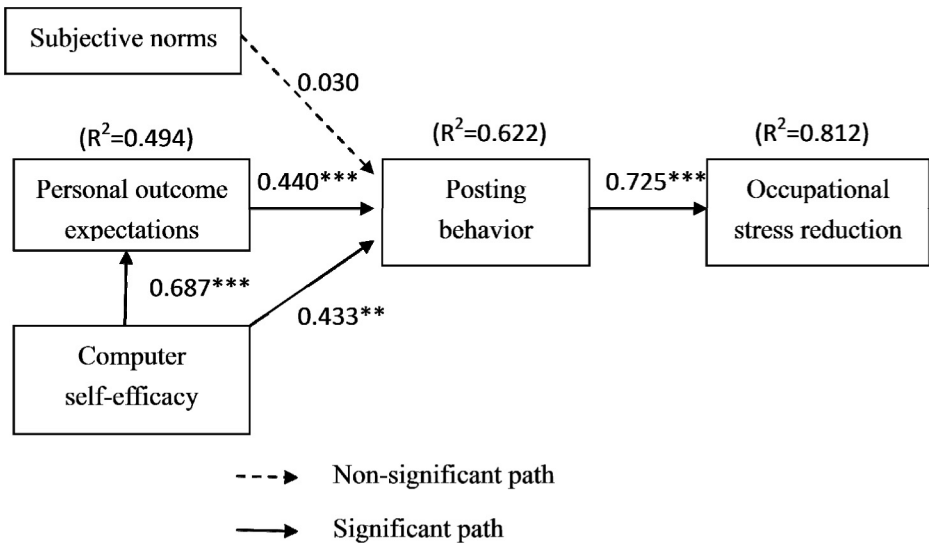

Figure 3 Results of Structural Modelling Analysis for Men

** means significant at the level of 0.01; *** means significant at the level of 0.001.

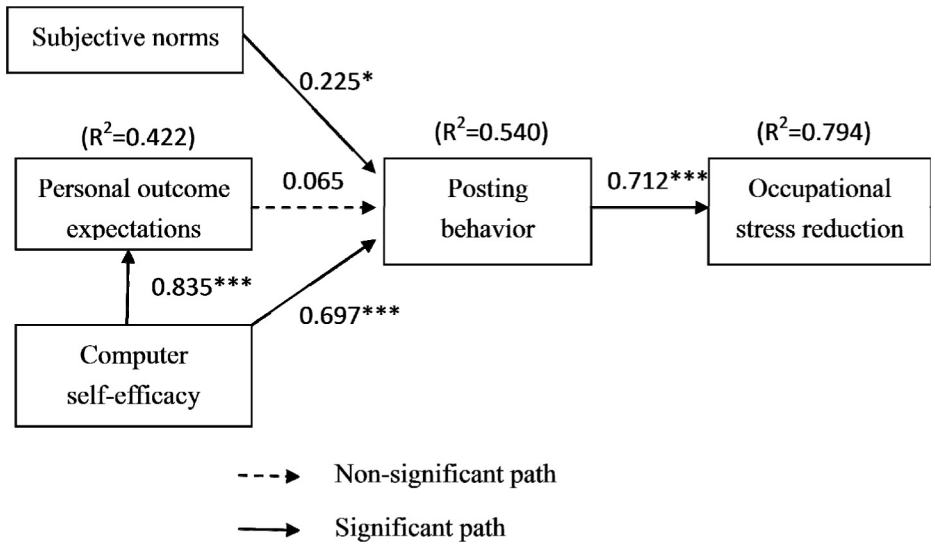

Figure 4 Results of Structural Modelling Analysis for Women

* means significant at the level of 0.05; *** means significant at the level of 0.001.

Table 7 Results of Estimated Structural Coefficients between Men and Women

Relationship	Estimate	SE	CR	*p*
Women (*N* = 139)				
Path 1: SN → PB	0.225	0.094	2.393	0.017*
Path 2: POE → PB	0.065	0.102	0.637	0.524
Path 3: CSE → POE	0.835	0.153	5.465	***
Path 4: CSE → PB	0.697	0.207	3.364	***
Path 5: PB → OSR	0.712	0.084	5.353	***
Men (*N* = 123)				
Path 1: SN → PB	0.030	0.112	0.269	0.788
Path 2: POE → PB	0.440	0.104	4.222	***
Path 3: CSE → POE	0.687	0.098	7.030	***
Path 4: CSE → PB	0.433	0.149	2.906	0.004**
Path 5: PB → OSR	0.725	0.066	7.590	***

Note: Estimate is unstandardized.

* means significant at the level of 0.05; ** means significant at the level of 0.01; *** means significant at the level of 0.001.

4.2 Discussion

The study has yielded several important findings. First, this study affirms that subjective norms, personal outcome expectations, and computer self-efficacy are positively associated with posting behavior, and computer self-efficacy is also positively associated with personal outcome expectations for occupational stress reduction. This result is in line with the study of Lu and Hsiao (2009) for frequent blog posting, but we also examined the effect of posting behavior on the status of occupational stress reduction. Subjective norms are important to the stressed workers for relaxing the stress through posting articles in SNS. Peer pressure will encourage users to do the posting behavior. SNS is a suitable outlet for venting emotions in that there are many active friends there. Subjective norms can make users likely comply with others' behavior, such as posting articles for stress reduction. That is, a user may likely post articles in SNS to mitigate occupational stress due to the norms from others. Moreover, personal outcome expectations are also important for users to post articles for occupational stress reduction. The higher the expected outcome is, the higher the probability of posting articles in SNS. That is, some people may likely post articles for occupational stress reduction because they expect the comfortable responses back from online friends, such as greetings, suggestions, and recognition. Thus, personal outcome expectations will positively affect posting behavior for occupational stress reduction in SNS. Furthermore, computer self-efficacy is another motivation to express feelings of stress to online friends for self-treatment. Maddux (1995) argued that if a person has higher self-efficacy in a particular job, this will make him/her more active participation in efforts to complete the task or the job. Thus, the higher a person's self-efficacy is, the higher the ability of self-efficacy to solve the problems (Shih, 2006). For example, when we are angry about something, we may likely tell to someone for venting the dissatisfaction. Computer self-efficacy not only can enhance outcome expectations, but also can strengthen posting behavior for occupational stress reduction.

Second, this study also finds that computer self-efficacy (Estimate = 0.541) and personal outcome expectations (Estimate = 0.219) affecting posting behavior are stronger than subjective norms affecting the behavior (Estimate = 0.152). Thus, posting articles in SNS for occupational stress reduction mainly result from individual's motivations, especially for computer self-efficacy. Obviously, most users posting articles in SNS for mitigating occupational stress may likely express themselves with the work-related experience to the friends. Like self-treatment, the users tend to post articles to online friends for self curing the stress. This action may not have any substantial rewards or obligation for the users. Relatively, personal outcome expectations and subjective norms play the second role of motivations to mitigate occupational stress for the users. This finding is not in line with the study of Lu and Hsiao (2009), which found that the effect of subjective norms on posting intention ($\beta = 0.3$) is stronger than the effects of personal

outcome expectations and self-efficacy on the intention ($\beta = 0.17$; $\beta = 0.17$). The reason is that bloggers who receive more subjective norms are likely to have more intention to publish frequently on blogs. Thus, we can conclude that a stressed user may likely post articles in SNS for mitigating occupational stress in terms of self-efficacy motivation, but a blogger posting articles for sharing knowledge may be subject to peer pressure. It is believed that the motivations to do the behavior will depend on the behavioral goals rather than the behavior itself (e.g., posting behavior).

Third, for testing the moderating effect of gender, our evidence reveals that the relationship between subjective norms and posting behavior is stronger for women than for men, but the relationship between personal outcome expectations and posting behavior is stronger for men than for women. This result is partially consistent with the study of Lu and Hsiao (2009), which found that the effect of subjective norms on posting behavior is no differences between men and women, but the effect of personal outcome expectations on posting behavior is significantly higher for men than for women, as well as the effect of self-efficacy on posting behavior is significantly higher for women than for men. In the present study, we found that subjective norms and self-efficacy are two main motivations for women to post articles for occupational stress reduction, whereas personal outcome expectations and self-efficacy are two main motivations for men. This result confirms the previous speculation that women are more social-oriented, but men are more task-oriented in dealing with stress. Therefore, it is believed that the motivations of posting articles in SNS for occupational stress reduction may differ by gender. That is, while mitigating occupational stress in SNS, men and women both likely have the self confidence to express their feelings of stress to the friends; additionally, men may also tend to have high expectations to the outcome from online friends by posting articles, but women may simultaneously comply with other's behavior to post articles.

5. Conclusion and suggestion

5.1 Conclusion

This study aims to explore posting behavior for occupational stress reduction in SNS. The contributions of this study start with a conceptual formulation of how subjective norms, personal outcome expectations, and computer self-efficacy affect posting behavior for occupational stress reduction in SNS. On this basis, the study examined the moderation of gender on the causal relationship of the motivations, posting behavior, and occupational stress reduction. An empirical study with 262 savvy Facebook users who have the experience of posting articles for occupational stress reduction affirms the causal relationship and clarified the ideas. The findings of this study could be helpful for

practitioners to provide social functions for stressed users to mitigate feelings of stress in SNS.

5.2 The implications for research and practice

This study provides both theoretical and practical benefits. From a research perspective, this study used SCT as the base model for exploring the effect of occupational stress reduction in SNS. Occupational stress reduction is predicted by posting behavior, and posting behavior is explained in terms of subjective norms, personal outcome expectations, and computer self-efficacy. This study affirmed the causal relationship of the motivations, posting behaviors, and occupational stress reduction for stressed users in SNS. Thus, this study extended the SCT model with the effect of occupational stress reduction in the context of SNS.

Moreover, this study also found that computer self-efficacy is the major determinant influencing posing behavior for occupational stress reduction. This result also supports Bandura's (1997) argument that self-efficacy is central to SCT. Computer self-efficacy is highly related to judgments of personal capability. Self-confidence to the capability of using computer is therefore, deemed to be tightly associated with posting behavior for occupational stress reduction in SNS.

From a practical perspective, this study found that computer self-efficacy is the primary motivation for users by posting articles to mitigate occupational stress. Thus, the practitioners shall develop social functions that help users to get closer connection with online friends easily. For example, Facebook launched "Checkin" service that allows people to use the GPS on their mobile phones to let friends know exactly where they are. People can utilize the service to express their status or feelings to friends for mitigating the stress.

Moreover, this study found that men have high outcome expectations by posting articles to reduce occupational stress in SNS more than women. Thus, the practitioners shall provide interactive functions that facilitate users to get the feedback from online friends more effectively. For example, Facebook offered "Chat" service that makes people to talk with online friends directly. On the other hand, the findings also revealed that women are more significant than men for posting articles by subjective norms. Thus, the practitioners shall provide sociable functions that effectively connect the persons who have the same interests. For example, Facebook provided "Improved friend lists" service that allows people to share a personal story with specific friend groups.

5.3 Limitation and suggestion of the study

This study contains some limitations. First, a bias may exist because of the convenient sampling through social networks in Facebook. Second, playing games on the Internet or mobile phones may be widely used as the other tool for venting the stress, but it is in a different way. Yee (2006) demonstrated that achievement, social, and immersion are the three motivations for play in online games. On the other hand, Reinecke (2009) found that work-related fatigue and exposure to daily hassles are both positively associated with the use of games for recovery. Thus, subsequent studies may extend the current framework of the study to explore the motivations substantially influence occupational stress reduction by playing online games. Third, this study used a cross-sectional data to analyze the causal relationships for occupational stress reduction in SNS. The results will only be inferred rather than proven (Fang, Chiu & Wang, 2011). Thus, subsequent studies may conduct a longitudinal approach to identify the dynamic change of the causal relationships for occupational stress reduction in SNS.

References

Anderson, J.C. and Gerbing, D.W. (1988), 'Structural equation modeling in practice: A review and recommended two-step approach', *Psychological Bulletin*, Vol. 103, No. 3, pp. 411-423.

Azjen, I. (1991), 'The theory of planned behaviour', *Organizational Behavior and Human Decision Processes*, Vol. 50, No. 2, pp. 179-211.

Bandura, A. (1986), *Social Foundations of Thought & Action: A Social Cognitive Theory*, Prentice-Hall, Englewood Cliffs, NJ.

Bandura, A. (1997), *Self-Efficacy: The Exercise of Control*, W.H. Freeman, New York, NY.

Battacherjee, A. (2000), 'Acceptance of e-commerce services: the case of electronic brokerages', *IEEE Transactions on Systems, Man, and Cybernetics -- Part A: Systems and Humans*, Vol. 30, No. 4, pp. 411-420.

Bock, C.W., Zmud, R.W., Kim, Y.G. and Lee, J.N. (2005), 'Behavioral intention formation in knowledge sharing: examining the roles of extrinsic motivators, social-psychological forces, and organizational climate', *MIS Quarterly*, Vol. 29, No. 1, pp. 87-111.

Burke, R.J. (1993), 'Work-family stress, conflict, coping, and burnout in police officers', *Stress Medicine*, Vol. 9, No. 3, pp. 171-180.

Burke, R.J. and Belcourt, M.L. (1974), 'Managerial role stress and coping responses', *Journal of Business Administration*, Vol. 5, No. 2, pp. 55-68.

Byrne, B.M. (2010), *Structural Equation Modeling with AMOS*, Routledge, New York, NY.

Cheung, G.W. and Rensvold, R.B. (2002), 'Evaluating goodness-of-fit indexes for testing measurement invariance', *Structural Equation Modeling: A Multidisciplinary Journal*, Vol. 9, No. 2, pp. 233-255.

Compeau, D.R., Higgins, C.A. and Huff, S. (1999), 'Social cognitive theory and individual reactions to computing technology: a longitudinal study', *MIS Quarterly*, Vol. 23, No. 2, pp. 145-158.

Cotton, P. and Hart, P.M. (2003), 'Occupational wellbeing and performance: a review of organizational health research', *Australian Psychologist*, Vol. 38, No. 2, pp. 118-127.

Davidson, M. and Cooper, C. (1983), *Stress and the Woman Manager*, Martin Robertson & Company, Oxford, UK.

Fang, Y.H., Chiu, C.M. and Wang, E.T.G. (2011), 'Understanding customers' satisfaction and repurchase intentions -- an integration of IS success model, trust, and justice', *Internet Research*, Vol. 21, No. 4, pp. 479-503.

Fishbein, M. and Ajzen, I. (1975), *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*, Addison-Wesley, Reading, MA.

Folkman, S. and Lazarus, R.S. (1988), 'An analysis of coping with stress', in Cooper, C.L. and Payne, R. (Eds.), *Causes, Coping and Consequences of Stress at Work*, John Wiley & Sons, Chichester, UK, pp. 223-263.

Fornell, C. and Larcker, D.F. (1981), 'Evaluating structural equations with unobservable variables and measurement error', *Journal of Marketing Research*, Vol. 18, No. 1, pp. 39-50.

French, J.R.P., Caplan, R.D. and Van Harrison, R. (1982), *The Mechanisms of Job Stress and Strain*, John Wiley & Sons, London, UK.

Ganster, D.C., Fusilier, M.R. and Mayes, B.T. (1986), 'Role of social support in the experience of stress at work', *Journal of Applied Psychology*, Vol. 71, No. 1, pp. 102-110.

Gardner, D. and Fletcher, R. (2009), 'Demands, appraisal, coping and outcomes: positive and negative aspects of occupational stress in veterinarians', *International Journal of Organizational Analysis*, Vol. 17, No. 4, pp. 268-284.

Gefen, D. and Straub, D. (2000), 'The relative importance of perceived ease of use in IS adoption: a study of e-commerce adoption', *Journal of Association of Information Systems*, Vol. 1, No. 1, Article 8.

George, J.F. (2004), 'The theory of planned behavior and Internet purchasing', *Internet Research*, Vol. 14, No. 3, pp. 198-212.

Herring, S.C. (1996), 'Two variants of an electronic message schema', in Herring, S.C. (Ed.), *Computer-Mediated Communication: Linguistic, Social, and Cross-Cultural*, John Benjamins, Philadelphia, PA, pp. 81-106.

Herring, S.C. and Paolillo, J.C. (2006), 'Gender and genre variation in weblogs', *Journal of Sociolinguistics*, Vol. 10, No. 4, pp. 439-459.

Hsu, C.L. and Lu, H.P. (2007), 'Consumer behavior in online game communities: a motivational factor perspective', *Computers in Human Behavior*, Vol. 23, No. 3, pp. 1642-1659.

Hsu, M.H., Ju, T.L., Yen, C.H. and Chang, C.M. (2007), 'Knowledge sharing behavior in virtual communities: the relationship between trust, self-efficacy, and outcome expectations', *International Journal of Human-Computer Studies*, Vol. 65, No. 2, pp. 153-169.

Huang, H.M. and Liaw, S.S. (2005), 'Exploring users' attitudes and intentions toward the web as a survey tool', *Computers in Human Behavior*, Vol. 21, No. 5, pp. 729-743.

Jung, T., Youn, H. and McClung, S. (2007), 'Motivations and self-presentation strategies on Korean-based Cyworld weblog format personal homepages', *CyberPsychology & Behavior*, Vol. 10, No. 1, pp. 24-31.

Kahn, R.L., Wolfe, D.M., Quinn, R.P., Snoek, J.D. and Rosenthal, R.A. (1964), *Organizational Stress: Studies in Role Conflict and Ambiguity*, John Wiley & Sons, New York, NY.

Kankanhalli, A., Tan, B.C.Y. and Wei, K.K. (2005), 'Contributing knowledge to electronic knowledge repositories: an empirical investigation', *MIS Quarterly*, Vol. 29, No. 1, pp. 113-143.

Karasek, R.A. (1979), 'Job demands, job decision latitude, and mental strain: Implications for job redesign', *Administrative Science Quarterly*, Vol. 24, No. 2, pp. 285-308.

King, M. and Gardner, D. (2006), 'Emotional intelligence and occupational stress among professional staff in New Zealand', *International Journal of Organizational Analysis*, Vol. 14, No. 3, pp. 186-203.

Lazarus, R.S. and Folkman, S. (1984), *Stress, Appraisal and Coping*, Springer, New York, NY.

Lee, M.K.O., Cheung, C.M.K., Lim, K.H. and Sia, C.L. (2006), 'Understanding customer knowledge sharing in web-based discussion boards: an exploratory study', *Internet Research*, Vol. 16, No. 3, pp. 289-303.

Lim, V.K.G. and Teo, T.S.H. (1996), 'Gender differences in occupational stress and coping strategies among IT personnel', *Women in Management Review*, Vol. 11, No. 1, pp. 20-28.

Lu, H.P. and Hsiao, K.L. (2007), 'Understanding intention to continuously share information on weblogs', *Internet Research*, Vol. 17, No. 4, pp. 345-361.

Lu, H.P. and Hsiao, K.L. (2009), 'Gender differences in reasons for frequent blog posting', *Online Information Review*, Vol. 33, No. 1, pp. 135-156.

Maddux, J.E. (1995), *Self-Efficacy, Adaptation, and Adjustment: Theory, Research, and Application*, Plenum Press, New York, NY.

Martocchio, J.L. and O'Leary, A.M. (1989), 'Sex differences in occupational stress: a meta-analytic review', *Journal of Applied Psychology*, Vol. 74, No. 3, pp. 495-501.

McCarty, W.P., Zhao, J.S. and Garland, B.E. (2007), 'Occupational stress and burnout between male and female police officers: are there any gender differences?', *Policing: An International Journal of Police Strategies & Management*, Vol. 30, No. 4, pp. 672-691.

McDonald, L.M. and Korabik, K. (1991), 'Sources of stress and ways of coping among male and female managers', *Journal of Social Behavior & Personality*, Vol. 6, No. 7, pp. 185-199.

Nelson, D. and Cooper, C. (2005), 'Stress and health: a positive direction', *Stress and Health*, Vol. 21, No. 2, pp. 73-75.

Nunnally, J.C. (1978), *Psychometric Theory*, 2nd ed., McGraw-Hill, New York, NY.

Reinecke, L. (2009), 'Games and recovery: the use of video and computer games to recuperate from stress and strain', *Journal of Media Psychology: Theories, Methods, and Applications*, Vol. 21, No. 3, pp. 126-142.

Robert, M.L. and Wortzel, L.H. (1979), 'New life-style determinants of women's food shopping behaviour', *Journal of Marketing*, Vol. 43, No. 3, pp. 28-39.

Shang, R.A., Chen, Y.C. and Shen, L. (2005), 'Extrinsic versus intrinsic motivations for consumers to shop on-line', *Information & Management*, Vol. 42, No. 3, pp. 401-413.

Shih, H.P. (2006), 'Assessing the effects of self-efficacy and competence on individual satisfaction with computer use: an IT student perspective', *Computers in Human Behavior*, Vol. 22, No. 6, pp. 1012-1026.

Simmons, B.L. and Nelson, D.L. (2001), 'Eustress at work: the relationship between hope and health in hospital nurses', *Health Care Management Review*, Vol. 26, No. 4, pp. 7-18.

Stowers, G.N.L. (1995), 'Getting left behind? Gender differences in computer conferencing', *Public Productivity & Management Review*, Vol. 19, No. 2, pp. 143-159.

Taylor, D. and Altman, I. (1987), 'Communication in interpersonal relationships: social penetration processes', in Roloff, M.E. and Miller, G.R. (Eds.), *Interpersonal Processes: New Directions in Communication Research*, Sage, Newbury Park, CA, pp. 257-277.

Trammell, K.D. and Keshelashvili, A. (2005), 'Examining the new influencers: a self-presentation study of A-list blogs', *Journalism and Mass Communication Quarterly*, Vol. 82, No. 4, pp. 968-982.

Trammell, K.D., Tarkowski, A., Hofmokl, J. and Sapp, A.M. (2006), 'Rzeczpospolita blogów [Republic of Blog]: examining Polish bloggers through content analysis', *Journal of Computer-Mediated Communication*, Vol. 11, No. 3, pp. 702-722.

Venkatesh, V. and Davis, F.D. (2000), 'A theoretical extension of the technology acceptance model: four longitudinal field studies', *Management Science*, Vol. 46, No. 2, pp. 186-204.

Venkatesh, V. and Morris, M.G. (2000), 'Why don't men ever stop to ask for directions? Gender, social influence, and their role in technology acceptance and usage behaviour', *MIS Quarterly*, Vol. 24, No. 1, pp. 115-139.

Wasko, M.M. and Faraj, S. (2005), 'Why should I share? examining social capital and knowledge contribution in electronic networks of practice', *MIS Quarterly*, Vol. 29, No. 1, pp. 35-57.

Wetzer, I.M., Zeelenberg, M. and Pieters, R. (2007), '"Never eat in that restaurant, I did!": exploring why people engage in negative word-of-mouth communication', *Psychology & Marketing*, Vol. 24, No. 8, pp. 661-680.

Wu, S.I. (2006), 'A comparison of the behavior of different customer clusters towards Internet bookstores', *Information & Management*, Vol. 43, No. 8, pp. 986-1001.

Yang, K. and Lee, H.J. (2010), 'Gender differences in using mobile data services: utilitarian and hedonic value approaches', *Journal of Research in Interactive Marketing*, Vol. 4, No. 2, pp. 142-156.

Yee, N. (2006), 'Motivations for play in online games', *CyberPsychology & Behavior*, Vol. 9, No. 6, pp. 772-775.

About the author

Yung-Shen Yen is Associate Professor of Computer Science and Information Management at Providence University, Taiwan. He obtained his Ph.D. in Business Administration from National Chengchi University, Taiwan. His research has been focusing on customer relationship in electronic commerce. He has published numerous articles on related journals, such as *Internet Research*, *The TQM Journal*, *Asia Pacific Journal of Marketing and Logistics*, *International Journal of Mobile Communications*, *International Journal of Computer and Information Technology*, *African Journal of Business Management*. Corresponding author. Department of Computer Science and Information Management, Providence University, 200 Chung-Chi Rd., Shalu, Taichung 43301, Taiwan. Tel: +886-4-26328001. E-mail address: ysyen@pu.edu.tw

Appendix
The instrument of the study

Subjective norms (SN) (adapted from Lu & Hsiao, 2009)

SN1 My friends expect me to post articles in my page of Facebook.

SN2 People who I contact expect me to post articles in my page of Facebook.

Personal outcome expectations (POE) (adapted from Lu & Hsiao, 2009)

POE1 I expect that I can receive greetings from my friends in my page of Facebook.

POE2 I expect that I can get the solutions of the problems from my friends in my page of Facebook.

POE3 I expect that I can improve others' recognition of me in my page of Facebook.

Computer self-efficacy (CSE) (adapted from Lu & Hsiao, 2009)

CSE1 I don't mind providing my personal interests or habits in my page of Facebook.

CSE2 I would like to tell others my feelings of my experience in my page of Facebook.

CSE3 I want to post articles in my page of Facebook to let others know me.

Posting behavior (PB) (adapted from Wu, 2006)

PB1 I have posted articles regarding work-related experiences via Facebook.

PB2 I often posted articles regarding work-related experiences in my page of Facebook.

Occupational stress reduction (OSR) (adapted from McCarty, Zhao & Garland, 2007)

OSR1 I do not feel tired at work when I used in my page of Facebook.

OSR2 I do not be moody, irritable, or impatient over small problems when I used in my page of Facebook.

OSR3 I withdraw from the constant demands on my time and energy from work when I used in my page of Facebook.

OSR4 I do not feel negative, futile or depressed about work when I used in my page of Facebook.

OSR5 I am as efficient at work as I should be when I used in my page of Facebook.

OSR6 Using in my page of Facebook heightened my resistance to illness because of my work.

OSR7 Using in my page of Facebook heightened my interest in doing fun activities because of my work.

OSR8 I have easily concentrating on my job when I used in my page of Facebook.

MIS Review Vol. 19, No. 1, September (2013), pp. 25-44
DOI: 10.6131/MISR.2013.1901.02

On Measuring and Increasing the Effectiveness of Banner Advertising

Haren Ghosh[1], Amit Bhatnagar[2]
[1]*President & CEO, Analytic Mix Inc., USA*
[2]*Sheldon B. Lubar School of Business, University of Wisconsin-Milwaukee, USA*

ABSTRACT: *Despite impressive growth in banner advertising budgets, doubts persist amongst practitioners about the effectiveness of banner advertising and the techniques used to measure it. A number of approaches such as click-throughs, mouse rollovers and eyetracker studies have been developed to measure the banner ad effectiveness. We argue that banner ad effectiveness can also be determined by measuring the change in perceptions of consumers who have been exposed to a banner ad. We further argue that the effectiveness of a banner ad can be increased by identifying the issues that are salient to the target consumers and then aligning the message in the banner ad with these issues. A case study is presented where the technique is demonstrated on an advertising campaign launched by the travel department of an Asian country. Consumers who were exposed to the banner ads were shown to be more likely to visit the advertised country.*

KEYWORDS: *Advertising Effectiveness, Ad Recall, Ad Message, Banner Advertising.*

1. Introduction

The news regarding online advertising continues to be very exciting. In the first half of 2013, online advertising revenues totaled $20.1 billion, exhibiting double digit annual growth rates for the past three years (Interactive Advertising Bureau, 2013). Online advertising growth has consistently outperformed total media market growth based on Nielsen and Kartar estimates (Interactive Advertising Bureau). Of all the different advertising formats that have been developed for the online domain, banner ads continues to be an important one. Banner ads are graphics that are placed at an ad hosting website and hyperlinked to the sponsoring website that carries more detailed product and promotional information. In the first-half of 2013, banner ad revenues were $6.1 billion, a 19% share of the online ad dollars (Please see Table 1).

The only dark cloud in all the exciting news about banner ads is clickthrough rate, defined as the percentage of consumers who are exposed to an ad and who actually click on the ad to get to the sponsoring website (Drèze & Hussherr, 2003; Fisher & Pappu, 2006). Clickthrough rate is an important metric that is frequently used to measure the

Table 1 Expenditure in Different Advertising Formats in the First-half of 2013

Ad Format	Share of Total Ad Revenue
Search	43%
Banner Advertising	19%
Mobile	15%
Online Classifieds	6%
Digital Video	7%
Lead Generation	4%
Sponsorship	2%
Rich Media	3%
Email	0.4%

Source: Interactive Advertising Bureau (2013).

effectiveness of banner ads. This metric is also used extensively to determine the price of banner ads. According to ComScore (www.comscore.com), over 5.3 trillion display ads were served to US users in 2012, with a typical Internet user served 1,707 banner ads per month. As the number of banner ads served per user has increased, the advertising clutter has also increased (Yaveroglu & Donthu, 2008). Consequently, clickthrough rates have plummeted and stand now at 0.1 percent (Chaffey, 2013). This statistic leads some researchers and practitioners to question whether banner ads are effective.

Some researchers argue that clickthrough rate is not a good measure of banner ad effectiveness. The argument is that even when consumers do not click on a banner ad, they may have paid attention to it. This has lead to studies that measure banner ad effectiveness by other means, such as mouse rollovers (Rosenkrans, 2010) and eyetracker analysis (Baraggioli & Brasel, 2008; Lee & Ahn, 2012), that track whether consumers are paying attention to an ad. When consumers pay attention to an ad, they process it peripherally, and get persuaded by the message. Consumers respond to banner ads in a number of different ways, and these responses are not all similar, but follow an hierarchy. There are different hierarchy models, but they all argue essentially that consumers' response to ads progress in an orderly fashion, where the first stage is cognitive ("thinking"), followed by affective ("feeling"), and finally conative ("doing") (Barry & Howard, 1990). We can measure consumer response in any fashion, cognitive, affective, or conative. We decided to measure consumer response by the first stage, that is, cognitive as it leads to all subsequent stages. We put forward a new method to measure banner ad effectiveness by directly ascertaining the changes in perceptions of consumers who have been exposed to a banner ad. Since consumer perceptions drive purchase probability, anything that changes these perceptions should also influence the purchase probability. By directly linking

exposure to a banner ad to purchase probability, our approach yields a very objective measure of banner ad effectiveness that can be used to determine the return on advertising investments.

A related issue is what drives the effectiveness of banner ads. Extant research has revealed that the effectiveness of banner ads depends on a number of factors, such as consumer characteristics (Palanisamy, 2005), executional elements (Yaveroglu & Donthu, 2008), and banner characteristics (Baraggioli & Brasel, 2008). One of the banner characteristics that is critical to the success of banner ads, but has not been studied so far is the advertising message. Since banner ads change the perceptions of consumers, the message in banner ads must play a critical role in how these perceptions change. This insight can help an advertiser in improving the effectiveness of banner ads. They should first identify the attributes that are salient to the target consumer, and then craft the message in the banner ads to change the perceptions of consumers about these attributes.

To demonstrate how to measure and increase the effectiveness of banner ads, we analyze an online advertisement campaign launched by the national tourism department of an Asian country. We chose the travel sector to demonstrate our strategy as advertising banners are used quite commonly in the travel industry. During March 2012, advertisers in the travel sector in UK placed nearly 3.1 billion display ad impressions, reaching 87 percent of the total internet audience and accounting for 4.5 percent of all display ad impressions served. This Asian country is an upcoming travel destination and their tourism department had launched a major advertising campaign on the Web consisting of banner ads. They wanted to determine if their current online advertising effort has been effective in increasing consumers' likelihood of visiting their country. More importantly, they wanted to identify a cost effective way to increase the effectiveness of their ad campaign. We show that the effectiveness of banner ads can be increased by changing the message in the banner ads.

In the next section, we survey the literature on advertising effectiveness to put the paper in its proper context. We provide details about the data in Section 3 and about the methodology used to analyze this data in Section 4. In Section 5, we discuss the results of our finding and conclude with managerial implications.

2. Literature review

In this section we first review the literature on advertising effectiveness in traditional media. We then review the literature on banner advertising effectiveness, followed by a review of the techniques used to measure banner advertising effectiveness. Last, we review the literature on how consumers choose travel destinations.

2.1 Traditional advertising effectiveness

Advertising effectiveness is one of the most important research issues in marketing. Advertisers constantly strive to implement a media resource allocation program that maximizes the return on media investment. Successful implementation of such a program requires a clear understanding of the role of media dollars in persuading the target segment to purchase the advertised product. Majority of these attempts revolve around media placement decision that relies primarily on reach (Pelsmacker, Geuens & Vermeir, 2004) and frequency (Naples, 1997). Although increasing reach and/or frequency can increase the effectiveness of a campaign, obtaining additional amounts of either or both can quickly become very expensive. This is because managers increase reach or frequency by either buying more impressions (Farris, Bendle, Pfeifer & Reibstein, 2006) for a longer period of time and/or employing multiple media.

While most of the existing research on advertising effectiveness has focused on reach and frequency, some researchers have also identified the important role of advertising message. Research has shown that advertising changes consumers' perceptions about the advertised product (Agostinelli & Grube 2002; Chang, 2002; Petty & Cacioppo, 1979; 1996; Shao, 2002). Wang (2006) shows that when consumers are engaged with the advertising message, the advertising effectiveness increases. Engagement with an ad can also be managed by the message strategy. For instance, Laskey, Fox and Crask (1995) show that different messages strategies lead to different levels of ad effectiveness, and the optimum message strategy depends on the product category. We extend their finding by arguing that the optimum message strategy will also depend on the market segment. It has been shown that marketing message should be different for men and women (Brunel & Nelson, 2003), different cultures (James & Hill, 1991). This is because different market segments seek different benefits. Therefore, the marketing message should depend on the benefits that the target consumers seek. Managers need to first determine whether their advertising message focuses on issues that are relevant to their target customers. If it is not, then the advertising message should be changed to bring it in alignment.

2.2 Banner ad effectiveness

Banner ad effectiveness depends on a number of factors, such as consumer characteristics (Palanisamy, 2005), executional elements (Yaveroglu & Donthu, 2008), and banner characteristics (Baraggioli & Brasel, 2008). Consumer characteristics, such as gender and culture, have been found to influence banner ad effectiveness. Palanisamy finds that gender influences the attitude towards banner ad, consumer expectations and banner ad effectiveness. In a cross-cultural study carried out in China and UK, Ju (2013) found culture to play a major role in banner ad effectiveness. Möller and Elsend (2010) find that consumers intention to click on banner ads can be explained by Hofstede's

cultural dimensions. Yaveroglu and Donthu focus more on execution strategy and conduct a number of experiments to show that advertising repetition strategy influences banner ad effectiveness. They found that banner ad repetition leads to greater brand recall ad and intention to click. In a noncompetitive environment, an ad variation strategy works better, whereas in a competitive environment, an ad repetition strategy works better. Other researchers have studied the role of ad size (Baltas, 2003), placement (Rosenkrans, 2010), and duration of exposure (Wang, Shih & Peracchio, 2013) on the effectiveness of banner ads.

The following studies have examined the effect of different aspects of banner ad design on banner ad effectiveness. Robinson, Wysocka and Hand (2007) study the impact of seven banner characteristics on ad effectiveness. The different design characteristics that they study are absence of promotional incentives, animation, presence of company logo, and action phrase. Baraggioli and Brasel (2008) use an eyetracker study to show that larger movements in wide spacing conditions can lead to increased visual attention on peripheral advertising banners. In a similar vein, Lee and Ahn (2012) use an eyetracker study to analyze the role of animation in banner ads in focusing consumer attention and subsequent information processing. They found that animation not only attracts less attention but also reduces the effect on memory. Thota, Song and Larsen (2010) also study the role of animation in banner ads effectiveness. Some other researchers have found animation to play no role in banner ad effectiveness (Robinson et al., 2007). As the evidence regarding animation in banner ads is mixed, Chtourou and Abida (2010) developed a typology of animations to enable better understanding of the effectiveness of different animation characteristics. Chi, Yeh and Chiou (2012) conduct a study to find that information presentation style in a banner ad influences its effectiveness. Rosenkrans (2010) study the role of banner ad design in effectiveness. More specifically, the design criterion they study are interactivity, animation, and nature of appeal (rational or emotional). One of the issues in banner ad design that has been overlooked by existing studies is the role of advertising message. From traditional advertising research, we know that ad message plays a major role in attracting consumer attention and motivating them to purchase the product. Briggs and Hollis (1997) have found that banner ads can change brand perceptions even without click-throughs. In this research, we study the role of advertising message and how it can be changed to maximize the effectiveness of banner ads.

2.3 Measures of banner ad effectiveness

The earliest and still the most widely used method of measuring brand ad effectiveness is derived from traditional advertising and is based on the number of people exposed to the banner ad. This measure has been criticized, as it is possible that

a consumer may have been exposed to a banner ad but may not have paid any attention to it (Lee & Ahn, 2012). Drèze and Hussherr (2003) find that even when consumers do not click on a banner ad, the banner ads are effective. Unlike traditional retailing environments, consumers' navigation behavior on the web can be recorded. This lead to the development of metrics that are more closely tied to consumers' interaction with banner ads, such as click-throughs (Baltas, 2003; Möller & Eisend, 2010) and mouse rollovers (Rosenkrans, 2010). However, these response based metrics are also problematic (Drèze & Hussherr). A couple of eyetracker studies (Baraggioli & Brasel, 2008; Lee & Ahn) have shown that even when consumers do not interact with an ad, they may pay attention to it and cognitively process the message in it. For a banner ad to be effective, consumers not only have to pay attention to it but also recall it after cognitive processing. Some of the factors that increase the recall of banner ads are embedded videos, price, product or services (Alijani, Mancuso, Kwun & Omar, 2010). A reasonable metric of brand ad effectiveness should be based on the fact that if consumers have paid attention to a banner ad and cognitively processed it, then their perceptions about the advertised product should have been altered. We put forward an approach that measures the effectiveness of banner ads by the extent to which consumer perceptions have been altered.

2.4 Choice of travel destination

The choice of a travel destination depends on the benefits that consumers seek, i.e., the attributes of travel destination (Kaciak & Louviere, 1990). Advertisers in tourism marketing have identified various benefits that motivate travelers in selecting their travel destinations (Baloglu & Uysal, 1996; Jamrozy & Uysal, 1994; Uysal & Hagan, 1993, etc.). Traditionally, the factors (e.g., beaches, recreation facilities, historic resources, etc.) have been taken as the drivers that attract travelers to a specific destination once the decision of travel is made (Baloglu & Uysal; Christensen, 1983; Crompton, 1979). The factors, which build the association of a brand with its perceived attributes, are instrumental in making effective marketing communications that influence travelers' decision choices about destinations (Baloglu & Uysal). This association of a brand and its perceived attributes has been accepted as a key to building a unique image when communicating a brand position by most advertising campaigns (Romaniuk & Gaillard, 2007). For instance, according to Romaniuk and Gaillard, associating a tourist spot with shopping opportunities builds the association between the tourist spot and shopping opportunities in consumers' mind. This will lead them to access a stronger link between the tourist spot and shopping at the expense of competitors.

A couple of papers have specifically studied the effect of advertising in the travel category and found that advertising changes consumers' perceptions about travel destinations (Greco 1988; Taylor & Franke, 2003). Grønhaug and Heide (1992) studied

the impact of a promotional film about Norway as a travel destination. They found that advertising changes consumers' perceptions about the target, and furthermore the created images may be different from their evaluations based on personal experience. It is therefore reasonable to assume that a target segment chooses a travel destination based on their perception of the destination on different benefit attributes. Their perception of the attributes is influenced by advertising, and therefore right choice of the message can influence the perceptions favorably.

3. Data and methodology

To evaluate the advertising campaign of the travel department of the destination country, data were collected with the help of an online survey. The advertising banner of the destination country was randomly displayed at a well-known travel website. The software dropped cookies on the computers of site visitors. When these visitors visited the site again, they were randomly chosen to participate in the online survey. Therefore, each respondent had visited the site twice. The first time they were exposed to the banner ad and second time they were invited to participate in the survey. The online data collection platform assigned a unique identification number to each of the respondents.

The data collection was carried out over a period of 72 hours. When we started displaying the banner ads, we recorded the time and exactly 72 hours later we stopped collecting the data. After the banner ad started getting displayed at the travel website, we started dropping cookies on the computers of all those who visited the travel site. We invited those visitors who returned to the site within 72 hours of the banner ad going live to participate in the study. Different respondents visited at different times and returned after different durations. For instance, someone might have returned after half an hour and someone else after 27 hours. It is also possible that someone returned after 72 hours, but they were not invited to participate in the study, because we stopped conducting the study after 72 hours. One of the challenges in the study was determining the duration of the study. If the duration is kept short, one does not get enough participants and the sample size becomes small. If the duration is long, the probability of the effect of the ad wearing off starts increasing. After some pre-trial studies, we determined 72 hours as the optimum time duration for the study. We carried out this study in conjunction with the media company of the travel department, and were therefore able to ensure that the banner ads were displayed at only this one travel site. This was necessary to prevent the participants getting exposed to banner ads at some other website.

In the survey, respondents were presented with a list of eight Asian countries that are well-known tourist destinations. The destination country is a south-east Asian

country that is gaining popularity as a tourist destination. The other seven countries were the neighboring countries of south-east Asia. Respondents were asked to state their likelihood of visiting each one of the countries on a 5 point Likert scale anchored by Very Unlikely and Very Likely. Respondents' preferences for the different countries were elicited upfront, so that their stated preferences for different countries were not contaminated by their exposure to the ad banner during the study. After determining respondents' likelihood of visiting different countries, they were asked to evaluate the focal country (hereafter referred as country **n**) on the following set of attributes, (1) has excellent shopping opportunities, (2) an exotic place to visit, (3) a romantic place to visit, (4) a relaxing place to visit, (5) a safe place to visit, (6) a clean place to visit, (7) a place to take kids on vacation, (8) has diverse culture, (9) traditions and history, (10) offers unique culinary adventures, (11) offers the opportunity to observe top class sporting events, (12) has beautiful scenery, (13) is an affordable country to travel to, (14) has beautiful beaches, (15) has exciting city attractions, and (16) is good for nightlife. The attributes were selected by identifying the key message goals of the advertising campaign. For each attribute, respondents were asked to state on a 5-point Likert scale, anchored by Strongly Disagree and Strongly Agree, whether country **n** has the attribute.

Finally, respondents were shown the online banner ad, and asked if they recall seeing the ad. The data collection procedure itself is quite unique, because it enabled us to show the actual advertising banner to every respondent. While several studies have used the Internet to conduct surveys, as far as we know this is one of the first studies to leverage the multimedia capability of the Internet to conduct the survey. Exposure to the original ad aided respondents' recall of the original ad. In addition, the survey contained general demographic information, which helped to profile the sample. In all, there were 860 complete responses with no missing data.

We next put forward a model to determine the role of banner advertising in the choice of a travel destination. In the first part of the model, we show how consumers' perceptions of a country influence the consumers' decision to travel to that country for vacation. In the second part, we show how advertising influences consumers' perceptions of the country.

The dependent variable is the probability of visiting country **n**, and the choice set consists of a list of countries that are rival travel destinations. In the survey, respondents were asked to state the likelihood of visiting each one of the countries in the choice set on a 5 point Likert scale, anchored by Not at all Likely and Very Likely. Let individual i's rating of likelihood of visiting country **n** on the Likert scale be P_i^n. Then, individual i's probability of visiting country **n** from the list of countries can be modeled by the standard multinomial logit probability (Maddala, 1983),

$$\pi_i^n = \frac{\exp(P_i^n)}{\sum_s \exp(P_i^s)} \quad (1)$$

where,

s indexes the countries in the choice set, and

P_t^s is respondent *i*'s rating of the likelihood of visiting country *s* on the Likert scale.

Our interest is in learning how a respondent's probability of visiting country **n**, i.e., π_i^n of Equation (1), is influenced by his/her perception of that country. This can be done by regressing π_i^n on consumer perceptions. However, we cannot run OLS regression, as the dependent variable π_i^n is constrained to vary between 0 and 1. We estimate the log odds ratio model, where we just transform the dependent variable to normalize it. Subsequently, we run the following regression equation to estimate the parameters,

$$\ln\left(\frac{\pi_i^n}{1-\pi_i^n}\right) = \alpha + \sum_j \beta_j X_{ij} + \varepsilon_i \quad (2)$$

where,

α is the base preference for country **n**,

X_{ij} is respondent *i*'s perception of country **n** on attribute *j*,

β_j is the role of *j* attribute in the choice of travel destination,

ε_i is the disturbance term, assumed to have a standard normal distribution.

Let $\overline{X}_j$ be the sample average value of attribute X_{ij}, and $\hat{\alpha}$ and $\hat{\beta}_j$ be the estimated values of α and β_j of Equation (2). By substituting $\overline{X}_j$, $\hat{\alpha}$, and $\hat{\beta}_j$ in Equation (2), we can determine the average probability of visiting country **n**, i.e., π_i^n by the sample.

$$\pi^n = \frac{\exp(\hat{\alpha} + \sum_j \hat{\beta}_j \overline{X}_j)}{1 + \exp(\hat{\alpha} + \sum_j \hat{\beta}_j \overline{X}_j)} \quad (3)$$

4. Data analysis and results

We split the total sample into two segments, those who recall the ad and those who do not. For each segment, we calculate their mean perception of the destination on each one of the sixteen attributes. We have plotted the mean values in Figure 1. From Figure 1, we can see that respondents who recall seeing the ads perceive country **n** to be higher on most of the attributes. To examine if there are any significant difference between the

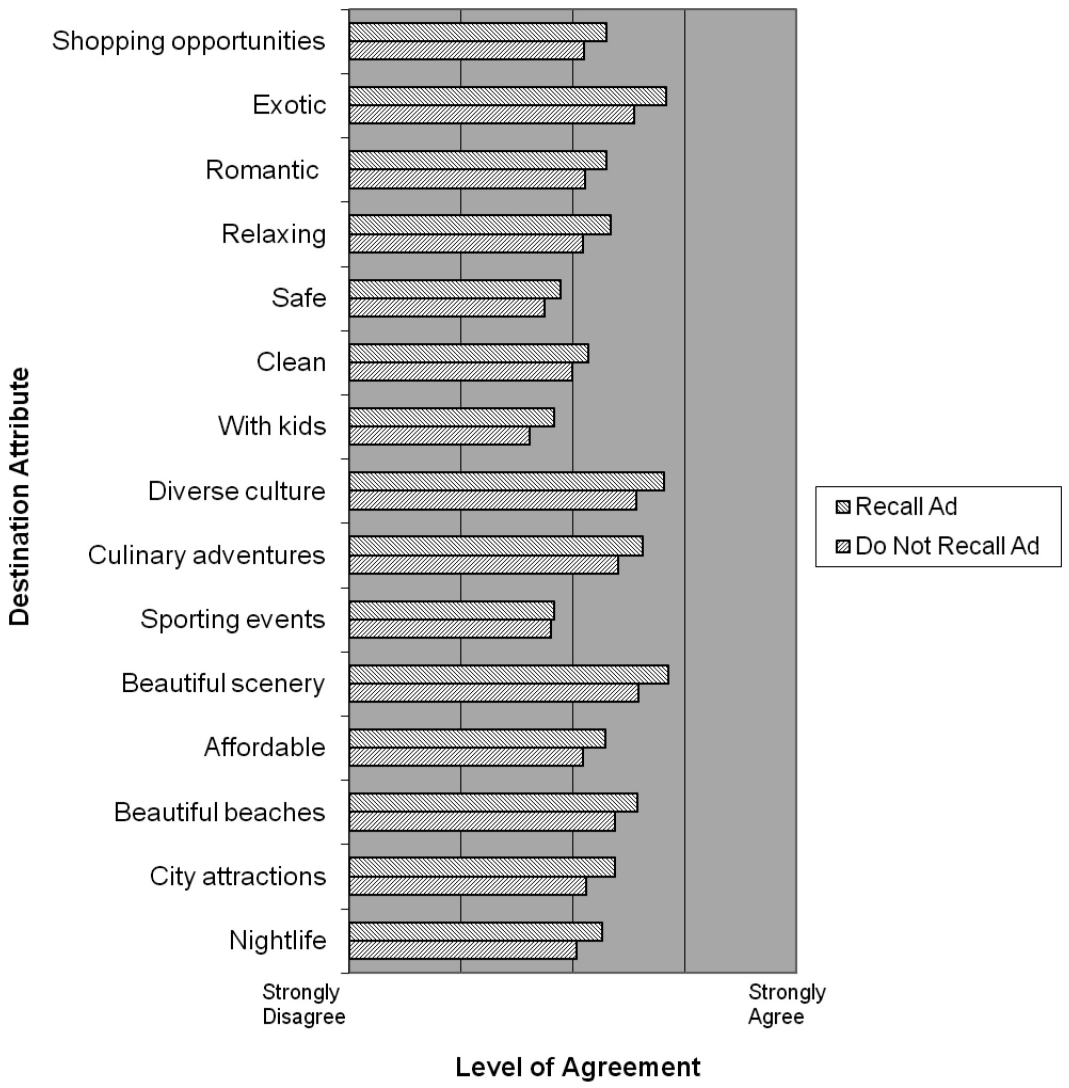

Figure 1 Respondents' Perception about the Target Destination

means of these two groups, a one-way ANOVA was run and the results are in Table 2. This shows that advertising does change consumers' perception about the country. The largest impact of advertising is on consumers' perception of the exotic nature of country **n**. The perception increased by 0.28 on a 5-point scale. The second highest impact is on consumers' perception of whether country **n** has beautiful scenery and natural environment. The impact of advertising in terms of decreasing order is on consumers' perceptions about diverse culture, traditions and history, and exciting city attractions. The advertising does not significantly change consumers' perception of two of the sixteen attributes, i.e., safe place to visit and offers the opportunity to observe top-class sporting events.

Table 2 Impact of Advertising on Perceptions of Travel Destination Attributes

	Mean (Do not Recall Ad)	Mean (Recall Ad)	F Stats	Difference between Means
Has excellent shopping opportunities	3.10 (0.785)	3.30 (0.918)	8.247**	0.20
An exotic place to visit	3.55 (0.986)	3.83 (1.002)	10.675**	0.28
A romantic place to visit	3.11 (0.891)	3.30 (0.980)	6.488**	0.19
A relaxing place to visit	3.09 (0.869)	3.34 (0.968)	10.706**	0.25
A safe place to visit	2.74 (0.957)	2.89 (1.014)	3.465	0.15
A clean place to visit	2.99 (0.822)	3.14 (0.920)	4.512*	0.15
A place to take kids on vacation	2.61 (0.861)	2.83 (0.935)	9.110**	0.22
Has diverse culture, traditions and history	3.56 (0.978)	3.81 (0.967)	9.311**	0.25
Offers unique culinary adventures	3.40 (0.882)	3.62 (0.907)	9.150**	0.22
Offers the opportunity to observe top-class sporting events	2.80 (0.706)	2.83 (0.791)	0.195	0.03
Has beautiful scenery / Natural environment	3.58 (0.963)	3.85 (0.959)	11.118**	0.27
Is an affordable country to travel to	3.09 (0.853)	3.29 (0.953)	7.470**	0.20
Has beautiful beaches	3.37 (0.882)	3.57 (0.912)	7.114**	0.20
Has exciting city attractions	3.12 (0.772)	3.37 (0.853)	13.227**	0.25
Is good for nightlife	3.03 (0.791)	3.26 (0.756)	11.624**	0.23

Note: The numbers in the bracket are the standard deviations. The number of respondents who recall the ad is 175 and the number of respondents who don't recall the ad is 685.
**Significant at 0.01 level; *Significant at 0.05 level.

To run the regression as specified in Equation (2), the ideal approach would be to run a multiple regression with all the perceptual attributes taken together as the independent variables. However, respondents' perceptions of the country on different attributes were very heavily correlated. This leads to the problem of multicollinearity, which is quite severe in our data. The traditional approach of handling multicollinearity is to combine the independent variables into fewer orthogonal factors. However, we did not wish to combine all the perceptual attributes, as we wanted to measure their individual impact on the choice decision. We, therefore, ran a series of simple regressions with all the attributes taken one by one.

The results of our analysis are reported in Table 3. The second column has estimates of the constant and the third column the coefficient of the country attribute (β_j). For each model, we tested the model fit by calculating the F statistics; the p-values are reported in the last column. As can be seen from the last column, the model fit for the attributes excellent shopping opportunities, exotic place, diverse culture, traditions and history, offers unique culinary adventures, has beautiful scenery, is an affordable place, has beautiful beaches did not fit the data. The coefficients for these attributes were not significantly different from zero. Therefore, changes in respondents' perceptions of these attributes will have no effect on the probability of visiting country **n**. The second column of Table 3 reveals that consumers' perception about whether a country is a safe place to visit has the maximum impact on respondents' probability of visiting country **n**. In terms of impact, the second most important attribute is whether a country is a place to take kids on vacation, which is followed in order of decreasing importance by a relaxing place, a clean place, a romantic place, good nightlife, exciting city attractions and an opportunity to witness top-class sporting events.

We next determine the probability of traveling to country **n** for the two segments, the segment that recalls the ad and the segment that doesn't. Let $\overline{X}_{jN}$ be the mean value of respondents' perception of country **n** on attribute *j* among those who do not recall the advertisement. Then among respondents who do not recall the banner ad, the average probability of visiting country **n** due to attribute *j*, would be,

$$\overline{\pi}_N^n = \frac{\exp(\hat{\alpha} + \hat{\beta}_j \overline{X}_{jN})}{1 + \exp(\hat{\alpha} + \hat{\beta}_j \overline{X}_{jN})} \tag{4}$$

Similarly, if $\overline{X}_{jA}$ is the mean value of respondents' perception of country **n** on attribute *j* among those who recall the ad, then the average probability of visiting country **n** due to attribute *j*, would be,

$$\overline{\pi}_A^n = \frac{\exp(\hat{\alpha} + \hat{\beta}_j \overline{X}_{jA})}{1 + \exp(\hat{\alpha} + \hat{\beta}_j \overline{X}_{jA})} \tag{5}$$

Table 3 Impact of Travel Destination Attributes on Travel Destination Choice

	Constant	Coefficient	F Stats	Sig
Has excellent shopping opportunities	-3.023** (0.144)	0.081 (0.044)	3.349	0.068
An exotic place to visit	-2.687** (0.137)	-0.022 (0.037)	0.366	0.546
A romantic place to visit	-3.334** (0.129)	0.180** (0.039)	20.877	0.000
A relaxing place to visit	-3.553** (0.130)	0.250** (0.040)	39.528	0.000
A safe place to visit	-3.803** (0.104)	0.374** (0.035)	112.276	0.000
A clean place to visit	-3.468** (0.133)	0.232** (0.042)	28.880	0.000
A place to take kids on vacation	-3.557** (0.112)	0.297** (0.040)	55.013	0.000
Has diverse culture, traditions and history	-2.642** (0.139)	-0.035 (0.037)	0.872	0.351
Offers unique culinary adventures	-2.688** (0.145)	-0.023 (0.041)	0.315	0.575
Offers the opportunity to observe top-class sporting events	-3.074** (0.145)	0.109* (0.050)	4.743	0.030
Has beautiful scenery / Natural environment	-2.780** (0.141)	0.003 (0.038)	0.008	0.929
Is an affordable country to travel to	-2.947** (0.135)	0.057 (0.041)	1.912	0.167
Has beautiful beaches	-2.881** (0.144)	0.033 (0.041)	0.662	0.416
Has exciting city attractions	-3.208** (0.149)	0.139** (0.046)	9.271	0.002
Is good for nightlife	-3.235** (0.146)	0.152** (0.046)	10.969	0.001

Note: The Numbers in the bracket are the standard errors.

**Significant at 0.01 level; *Significant at 0.05 level.

The results of the analysis are reported in Table 4. The first column in Table 4 has the probability of visiting the target country for the respondents who do not recall the ad and the second column has the probability of visiting for those who do. The probability of visiting country **n** is higher among the respondents who recall the advertisement. The third column has the percentage increase in probability due to advertising. Since only

Table 4 Impact of Advertising on Perceptions of Travel Destination Attributes

	Likelihood of Visiting (Do not Recall Ad)	Likelihood of visiting (recall ad)	% Increase in Likelihood of Visiting (Everyone Exposed to Advertising)	Likelihood of Visiting (With Proposed Strategy)	% Increase in Likelihood of Visiting (With Proposed Strategy)
Has excellent shopping opportunities	0.046	0.046	0.000	0.046	0.00
An exotic place to visit	0.064	0.064	0.000	0.064	0.00
A romantic place to visit	0.059	0.061	3.165	0.062	4.85
A relaxing place to visit	0.058	0.062	5.705	0.062	6.80
A safe place to visit	0.059	0.062	5.136	0.065	10.33
A clean place to visit	0.059	0.061	3.219	0.062	6.29
A place to take kids on vacation	0.058	0.062	5.956	0.063	8.12
Has diverse culture, traditions and history	0.066	0.066	0.000	0.066	0.00
Offers unique culinary adventures	0.064	0.064	0.000	0.064	0.00
Offers the opportunity to observe top-class sporting events	0.059	0.059	0.307	0.061	2.91
Has beautiful scenery / Natural environment	0.058	0.058	0.000	0.058	0.00
Is an affordable country to travel to	0.050	0.050	0.000	0.050	0.00
Has beautiful beaches	0.053	0.053	0.000	0.053	0.00
Has exciting city attractions	0.059	0.061	3.215	0.061	3.73
Is good for nightlife	0.059	0.061	3.234	0.061	4.08

25.5% of the sample recalls the ad, the obvious recommendation would be to increase the percentage of respondents who recall the advertisement. This is normally achieved by increasing the ad reach and/or frequency, both of which are expensive propositions.

5. Managerial implications

According to Table 2, respondents who recall the ad perceive the shopping opportunities to be higher in country **n**. However, according to Table 3, respondents' choice of country **n** as a travel destination is not influenced by the shopping opportunities in country **n**. Further examination of Table 3 shows that while choice of country **n** is not influenced by shopping opportunities, respondents are more likely to visit country **n** if they perceive it to be a romantic place. From Table 2, we find that advertising actually increased respondents' perception that the target country is romantic. Therefore, while advertising increases respondents' perceptions of both shopping opportunities and romantic place by the same factor (i.e., 0.2), only perception of romantic place is relevant in determining the likelihood of visit.

Managerial implications are fairly obvious. The advertising message should focus on changing consumers' perceptions of those attributes that are relevant to the target customer. Data regarding consumers' perception of a country can be easily elicited through a survey, where the participants are asked to rate the country on a set of attributes. They can also be asked to rate how important each one of the attributes is to them. This data can be then used to learn what attributes are important to the customers and how the country rates on those attributes. Banner ads can be then used to change or reinforce consumers' perceptions. If perception of the country as a romantic place is more important, then the advertising message should try to persuade consumers that the destination is a romantic place. If perceptions of certain attributes are not relevant to the consumers, then advertising message should not focus on those. From Table 2, we find that advertising has made the maximum impact on respondents' perception of the country as an exotic place; it has gone up by a factor of 0.28 on a 5-point scale. If the advertising message had been able to improve respondents' perceptions of other attributes by the same magnitude, then the probability of visiting country **n** would have gone up significantly. This scenario analysis has been examined for each attribute; the new probabilities and percentage increases are reported in columns 4 & 5 of Table 4. The factor that produces maximum increase in the probability of visiting country **n** is the perception of the country as a safe place to visit, followed by a place to take kids on vacation.

Until now, the advertising focus has been on persuading respondents that country **n** is an exotic place with lots of beautiful scenery, exciting city attractions, etc. However,

these attributes are not drivers of country **n**'s choice as a travel destination for the target consumers. Respondents are more concerned about whether the country is a safe place where they can take their kids on a vacation. We have plotted the mean values of Table 2 in Figure 1, which shows that even the respondents who do not recall the ad have very high perceptions about whether the country is exotic, has diverse culture, has beautiful scenery, etc. On the other hand, respondents have very poor image of country **n** on some of the key drivers such as whether it is a safe place, a place to take kids, and a clean place. Based on these findings, advertisers should try to improve the country's image on key drivers (e.g., safety, place for family with kids, etc.), which will improve the likelihood of visiting country **n**. From Table 4, we see that if everyone can recall the message, the maximum increase in market share would be 5.956%. Achieving a 100% recall rate is not only nearly impossible, but also would be extremely expensive. On the other hand, if we just change the advertising message to change respondents' perception about safety, we can increase the market share by 10.33%. This can be achieved with the current level of advertising expenditures.

By understanding the impact of advertising message on consumers' purchase likelihood and market share, media planners and advertisers can focus on a cost effective strategy that can improve advertising effectiveness. The study demonstrates that just by reaching consumers through an advertisement and exposing the message multiple times will not help gain larger market share, unless the message stimulates consumers' emotion. The study also suggests the identification of the key drivers that arouse consumers' emotion and recommends a positive relationship between those drivers and the advertising message. However, there are some limitations to this study that offer opportunities for further research. First, the study has examined a particular product, i.e., tourism destination, where the verbal message and visual format have influential roles (Dann, 1996). The approach of this study can be applied to other product categories to generalize the findings. Second, the study examines the effectiveness of advertising message only for a single media channel; a study combining multiple media carrying an integrated message, especially a combination of online and offline media would be a very useful tool for planning an integrated marketing communication strategy. We also do not control for any moderating variables. The effectiveness of banner advertising depends on a number of factors and their effects needs to be controlled for in any analysis. Future researchers should collect data on all determinants of online advertising and allow for their moderating influence on the determinants included in this study.

References

Agostinelli, G. and Grube, J.W. (2002), 'Alcohol counter-advertising and the media. A review of recent research', *Alcohol Research & Health*, Vol. 26, No. 1, pp. 15-21.

Alijani, G.S., Mancuso, L.C., Kwun, O. and Omar, A. (2010), 'Effectiveness of online advertisement factors in recalling a product', *Academy of Marketing Studies Journal*, Vol. 14, No. 1, pp. 1-10.

Baloglu, S. and Uysal, M. (1996), 'Market segments of push and pull motivations: a canonical correlation approach', *International Journal of Contemporary Hospitality Management*, Vol. 8, No. 3, pp. 32-38.

Baltas, G. (2003), 'Determinants of internet advertising effectiveness: an empirical study', *International Journal of Market Research*, Vol. 45, No. 4, pp. 505-513.

Baraggioli, F. and Brasel, S.A. (2008), 'Visual velocity: content font effects and incidental online ad exposure', *Advances in Consumer Research*, Vol. 35, pp. 600-606.

Barry, T.E. and Howard, D.J. (1990), 'A review and critique of the hierarchy of effects in advertising', *International Journal of Advertising*, Vol. 9, No. 2, pp. 121-135.

Briggs, R. and Hollis, N. (1997), 'Advertising on the web: is there response before click-through?', *Journal of Advertising Research*, Vol. 37, No. 2, pp. 33-46.

Brunel, F.F. and Nelson, M.R. (2003), 'Message order effects and gender differences in advertising persuasion,' *Journal of Advertising Research*, Vol. 43, No. 3, pp. 330-341.

Chaffey, D. (2013), 'Display advertising clickthrough rates,' available at http://www.smartinsights.com/internet-advertising/internet-advertising-analytics/display-advertising-clickthrough-rates/ (accessed 12 May 2014).

Chang, C. (2002), 'Self-congruency as a cue in different advertising-processing contexts,' *Communication Research*, Vol. 29, No. 5, pp. 503-536.

Chi, H.K., Yeh, H.R. and Chiou, C.Y. (2012), 'The mediation effect of information presentation style on the relationship between banner advertisements and advertising effectiveness,' *International Journal of Business and Management*, Vol. 7, No. 14, pp. 46-52.

Christensen, J.E. (1983), 'An exposition of canonical correlation in leisure research,' *Journal of Leisure Research*, Vol. 15, No. 4, pp. 311-322.

Chtourou, M.S. and Abida, F.C. (2010), 'What makes one animation more effective than another? an exploratory study of the characteristics and effects of animation in internet

advertising,' *International Journal of Internet Marketing and Advertising*, Vol. 6, No. 2, pp. 107-126.

Crompton, J.L. (1979), 'Motivations for pleasure vacation', *Annals of Tourism Research*, Vol. 6, No. 4, pp. 408-424.

Dann, G.M.S. (1996), *The Language of Tourism: A Sociolinguistic Perspective*, CAB International, Wallingford, UK.

Drèze, X. and Hussherr, F.X. (2003), 'Internet advertising: is anybody watching?', *Journal of Interactive Marketing*, Vol. 17, No. 4, pp. 8-23.

Farris, P.W., Bendle, N.T., Pfeifer, P.E. and Reibstein, D.J. (2006), *Marketing Metrics: 50+ Metrics Every Executive Should Master*, Prentice Hall, Upper Saddle River, NJ.

Fisher, J. and Pappu, R. (2006), 'Cyber-rigging click-through rates: exploring the ethical dimensions', *International Journal of Internet Marketing and Advertising*, Vol. 3, No. 1, pp. 48-59.

Greco, A.J. (1988), 'The elderly as communicators: perceptions of advertising practitioners', *Journal of Advertising Research*, Vol. 28, No. 3, pp. 39-46.

Grønhaug, K. and Heide, M. (1992), 'Stereotyping in country advertising: an experimental study', *European Journal of Marketing*, Vol. 26, No. 5, pp. 56-66.

Interactive Advertising Bureau. (2013), 'IAB internet advertising revenue report: 2013 first six months' results', available at http://www.iab.net/media/file/IABInternetAdvertisingRevenueReportHY2013FINALdoc.pdf (accessed 7 May 2013).

James, W.L. and Hill, J.S. (1991), 'International advertising messages: to adapt or not to adapt', *Journal of Advertising Research*, Vol. 31, No. 3, pp. 65-71.

Jamrozy, U. and Uysal, M. (1994), 'Travel motivation variations of overseas German visitors', *Journal of International Consumer Marketing*, Vol. 6, No. 3/4, pp. 135-160.

Ju, B. (2013), 'A proposed cross-cultural examination of online advertising effectiveness in china and the UK', *International Journal of Business and Management*, Vol. 8, No. 6, pp. 34-39.

Kaciak, E. and Louviere, J. (1990), 'Multiple correspondence analysis of multiple choice experiment data', *Journal of Marketing Research*, Vol. 27, No. 4, pp. 455-465.

Laskey, H.A., Fox, R.J. and Crask, M.R. (1995), 'The relationship between advertising message strategy and television commercial effectiveness', *Journal of Advertising Research*, Vol. 35, No. 2, pp. 31-39.

Lee, J.W. and Ahn, J.H. (2012), 'Attention to banner ads and their effectiveness: an eye-tracking approach', *International Journal of Electronic Commerce*, Vol. 17, No. 1, pp. 119-138.

Maddala, G.S. (1983), *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, Cambridge, UK.

Möller, J. and Eisend, M. (2010), 'A global investigation into the cultural and individual antecedents of banner advertising effectiveness', *Journal of International Marketing*, Vol. 18, No. 2, pp. 80-98.

Naples, M.J. (1997), 'Effective frequency: then and now', *Journal of Advertising Research*, Vol. 37, No. 4, pp. 7-12.

Palanisamy, R. (2004), 'Impact of gender differences on online consumer characteristics on web-based banner advertising effectiveness', *Journal of Services Research*, Vol. 4, No. 2, pp. 45-55.

Pelsmacker, P., Geuens, M. and Vermeir, I. (2004), 'The importance of media planning, ad likeability and brand position for ad and brand recognition in radio spots', *International Journal of Market Research*, Vol. 46, No. 4, pp. 465-477.

Petty, R.E. and Cacioppo, J.T. (1979), 'Effects of message repetition and position on cognitive responses, recall and persuasion', *Journal of personality and Social Psychology*, Vol. 37, No. 1, pp. 97-109.

Petty, R.E. and Cacioppo, J.T. (1996), *Attitudes and Persuasion: Classic and Contemporary Approaches*, Westview Press, Boulder, CO.

Robinson, H., Wysocka, A. and Hand, C. (2007), 'Internet advertising effectiveness: the effect of design on click-through rates for banner ads', *International Journal of Advertising*, Vol. 26, No. 4, pp. 527-541.

Romaniuk, J. and Gaillard, E. (2007), 'The relationship between unique brand associations, brand usage and brand performance: analysis across eight categories,' *Journal of Marketing Management*, Vol. 23, No. 3-4, pp. 267-284.

Rosenkrans, G. (2010), 'Maximizing user interactivity through banner ad design', *Journal of Promotion Management*, Vol. 16, No. 3, pp. 265.

Shao, A.T. (2002), 'Nonconformity advertising to teens', *Journal of Advertising Research*, Vol. 42, No. 3, pp. 56-65.

Taylor, C.R. and Franke, G.R. (2003), 'Business perceptions of the role of billboards in the US economy', *Journal of Advertising Research*, Vol. 43, No. 2, pp. 105-161.

Thota, S.C., Song, J.H. and Larsen, V. (2010), 'Do animated banner ads hurt websites? The moderating roles of website loyalty and need for cognition', *Academy of Marketing Studies Journal*, Vol. 14, No. 1, pp. 91-116.

Uysal, M. and Hagan, L.A.R. (1993), 'Motivation of pleasure travel and tourism', in Khan, M.A., Olsen, M.D. and Var, T. (Eds.), *VNR's Encyclopedia of Hospitality and Tourism*, Van Nostrand Reinhold, New York, pp. 798-810.

Wang, A. (2006), 'Advertising engagement: a driver of message involvement on message effects', *Journal of Advertising Research*, Vol. 46, No. 4, pp. 355-368.

Wang, K.Y., Shih, E. and Peracchio, L.A. (2013), 'How banner ads can be effective: investigating the influences of exposure duration and banner ad complexity', *International Journal of Advertising*, Vol. 32, No. 1, pp. 121-141.

Yaveroglu, I. and Donthu, N. (2008), 'Advertising repetition and placement issues in on-line environments', *Journal of Advertising*, Vol. 37, No. 2, pp. 31-43.

About the authors

Haren Ghosh is the President and CEO at Analytic Mix Inc. He has a decade long quantitative research and entrepreneurial experience. He has served in executive positions in three different companies, including Chief Analytics Officer at Symphony Health Solutions, GM & Chief Analytics Officer at Symphony Advanced Media, and SVP of Marketing Sciences & CMO at Factor TG. Haren received an MBA in Finance and Marketing from Louisiana Tech University and his PhD studies in Marketing, with a focus in Applied Statistics and Advanced Econometrics from Louisiana State University. He received an undergraduate degree in Mechanical Engineering from Jalpaiguri Government Engineering College in India. He was a visiting Assistant Professor of Marketing at South Eastern Louisiana University before joining Factor TG. E-mail address: Haren.Ghosh@AnalyticMix.com

Amit Bhatnagar has a PhD in Marketing from SUNY -- Buffalo, MS in Aerospace Engineering and BS in Mechanical Engineering from IIT Kanpur. Dr. Bhatnagar has published papers in the *Journal of Business*, *Marketing Letters*, *Journal of Advertising*, *Journal of Business Research*, *Journal of Retailing*, *European Journal of Operations Research*, *International Journal of Research in Marketing*, etc. He has served as a reviewer for the *Marketing Science*, *Marketing Letters*, *Journal of Retailing*, *Journal of Business Research*, etc.

Corresponding author. Sheldon B. Lubar School of Business, University of Wisconsin-Milwaukee, Milwaukee, WI 53201, USA. Tel: +1-414-229-2520. E-mail address: amit@uwm.edu

MIS Review Vol. 19, No. 1, September (2013), pp. 45-60
DOI: 10.6131/MISR.2013.1901.03

Securing DNA Information through Public Key Cryptography

Shiv P. N. Tripathi, Manas Jaiswal, Vrijendra Singh
Division of MS (Cyber Law & Information Security), Indian Institute of Information Technology, India

ABSTRACT: *This research work emerged as a new concept to provide robust security to the huge volume of information residing in DNA. In present scenario, security is being managed through symmetric key cryptography only. A new initiative has been taken to increase the robustness of DNA security. In this paper we are integrating public key cryptography inside traditional DNA security algorithm. The additional security is provided through a new algorithm as proposed, which takes advantage of residue theorem and traditional RSA algorithm. The main security concept is based on complexity in factorization and high versatility of choosing parameters/ variables. Basically, DNA is encrypted through symmetric key cryptography and the key used to encrypt the data symmetrically is itself encrypted asymmetrically through proposed modified RSA algorithm. Through example, it is further illustrated in this paper that this is not only one of the optimized algorithms to provide a tradeoff between security and computational speed but also adds some sort of defense strategy against various attacks in a layered approach.*

KEYWORDS: *Information Security, Cryptography, DNA, DNA security, Encryption, RSA Algorithm.*

1. Introduction

Information plays a vital role in today's generation and this information is very extensive and huge from the storage perspective. The computing methodology to get decisive result, DNA sequences must be used because it is the only source for "ultra-compact information storage" (Huge data stored in a compact volume). In a gram of DNA, 108 tera-bytes storage capacity is available (Gehani, LaBean & Reif, 1999) Total data present in the world can be stored in a few grams of DNA. Now a question arises "why there is a lag in adopting this technology for storage and transmission purpose?"

The answer lies in computational complexity and security of information transmitted. A lot of research has been done to increase the computational speed using vast parallelism (Cox, 2001; Cui, 2006). But the second aspect of DNA security is equally challenging and thirst area to be taken care of. The strands of DNA are used to encrypt the original data. In this technique data is using DNA. A literature survey has been done and it is found that generally either we are compromising with storage capacity, computational speed or security of data.

Now, here is a proposed hybrid security concept in which we combine the traditional approach with a new approach (DNA Cryptography). The extra advantage of this method is that it provides two levels of security (first is achieved by encrypting data symmetrically and second is through encrypting the key itself asymmetrically which is used to encrypt the data symmetrically). In this new algorithm we are using a new improved version of RSA that makes it too strong to break. It decreases its factorization possibility (the only way to crack RSA). The asymmetric key cryptography begins from a password / passphrase / key (that is used to encrypt data [achieved from the DNA sequence]) and this mechanism is introduced to enhance the security so that it could lead towards robustness in comparison to the traditional One Time Pad DNA, PCR, Index based symmetric encryption and asymmetric (digital signature method) (Cherian, Raj & Abraham, 2013).

2. Review of related literature

In a pioneering study, a lot of magnificent research has been done to encrypt and decrypt the information residing in DNA (Clelland, Risca & Bancroft, 1999; Leier, Richter, Banzhaf & Rauhe, 2000; Shimanovsky, Feng & Potkonjak, 2002; Youssef, Emam & Abd Elghany, 2012). A DNA sequence is formed with four of alphabets: A, T, G and C. Every alphabet among these four is a reference for a nucleotide. The two important factors for this sequence are shown:

(1) A real DNA is almost similar to a fake one.

(2) A large number of DNA sequences are openly available on the internet (EMBL-EBI, 2012). On the basis of referenced previous researches, we can get approximately 55 million DNA sequences openly available (EMBL-EBI; Youssef et al., 2012).

In this world, all the living creatures have their own unique genetic specification, which is formed by two strands of nucleotides, consisting one out of four alphabets as discussed above A, T, G, C. Apart from this formation of DNA, it has some sort of chemical polarization property, it clearly indicates that every molecule contains different chemical groups (Schena, 2003).

This was Adleman, with his outstanding work (Adleman, 1994), brought a revolution in this magnificent area of bio-computation research. This research of bio-computation helps to solve all the problems which were supposed either impractical or almost impossible to solve due to lack of tremendous amount of processing. By the use of exceptional and efficient DNA computation technology, it could have been possible to break Data Encryption Standard (DES) (Boneh, Dunworth & Lipton, 1995). The OTP cryptography was achieved through DNA strands and another milestone in this field was steganography, discussed in (Gehani et al., 1999).

Following the work of Gehani (Gehani et al., 1999), the pseudo cryptography was introduced by Ning Kang (Ning, 2009). In this technique, the original data is transformed into a DNA sequence. The resulting sequence is then transformed into two different forms of DNA. (1) Spliced form and (2) Protein form. To achieve this, introns are divided into specific patterns. In the above technique, instead of using an actual DNA sequence, DNA functions are used. Thus it is called Pseudo DNA cryptography. The basic idea of central dogma of molecular biology like Transcription, Translation and Splicing are used to replicate this technique.

With the rapid development of DNA cryptography, some biological and algebraic operations based on DNA sequence are presented by researchers (Mills, Yurke & Platzman, 1999; Soni & Johar, 2012; Wasiewicz, Mulawka, Rudnichi & Lesyng, 2000) such as addition or subtraction.

Using A, T, G and C strands of DNA several binary encryption techniques and java crypto encryption techniques are proposed (Leier et al., 2000; Tatiana, Mircea-Florin, Monica & Cosmin, 2008)

In another research, primers are employed as key for encrypting and decrypting data which will result in a DNA template (Cherian et al., 2013). The basic technology used in this scheme is Polymerase Chain Reaction (PCR) which is a DNA digital coding technique (Cui, Qin, Wang & Zhang, 2008). In this methodology input data is changed into hexadecimal values which will be converted further in to binary code. These binary digits are used to convert DNA sequence in to DNA template. PCR processes are executed (Clelland et al., 1999) by using the forward primer. Now that the DNA sequences has been changed through the above process, so it will be completely different from the original data. In the decryption process of achieved encoded data, we use reverse primer to get back original DNA sequences. We calculate the equivalent binary digits and transformed it into the original data.

Now, most of the DNA cryptography is done using only symmetric key schemes. An asymmetric encryption method was proposed by (Lai, Lu, Qin, Han & Fang, 2010). In this, two keys are used, one is for encryption and decryption and other is for creating signatures. LU MingXin using advanced cryptographic techniques proposed symmetric key crypto-system (Lu, Lai, Xiao & Qin, 2007), in which specially designed micro arrays were used and DNA fabrication and DNA hybridization is done for encryption and decryption.

3. Proposed methodology

DNA information algorithm proposes a hybrid cryptography technique (symmetric + asymmetric) (Terec, Vaida, Alboaie & Chiorean, 2011) to secure the data being transmitted via DNA. This uses the public as well as private key cryptographic methods to provide an additional level of security to encrypted data.

First, at the beginning, the sender and corresponding receiver will generate a pair of asymmetric keys. Second, a negotiation for the use of symmetric algorithm is done with the codon sequence (a particular sequence or arrangement in a logical order or a fixed pattern) to be used on the basis of DNA indexes. Third, there is an assumption in the message transfer that the data is encrypted with symmetric algorithm, while the key itself is encrypted asymmetrically and attached with the data. This type of double layered security approach was first given by (Nobelis, Boudaoud & Riveill, 2008). Given below is a detailed description of DNA algorithm (Generation, Encryption, Transmission and Decryption).

3.1 Generation of key pairs

In the beginning of generation phase, a password / password phrase (it could be complex enough based on the maturity of security level that has to be provided to the information residing in DNA). Basically the length and strength of password is totally dependent on the criticality of the data that has to be secured. After the initial hashing of the password and using improved RSA algorithm (discussed below) two keys are generated, the asymmetric key pairs generated will be different every time even if the user enters the same password as it is based on the randomness of pseudo prime numbers p & q.

3.2 Asymmetric key generation

In this section we are showing the methodology adopted (Wang, Chen & Duan, 2011) (improved version of RSA) for obtaining encoded key of DNA message. Basically this new algorithm adopts the basic idea of RSA algorithm but here the complexity is increased and made it very secure from the existing attack of RSA algorithm.

The main aim was to come up with an improved algorithm that will make us very comfortable from major attacks, key management, or key storage perspective.

Step 1: first the password / passphrase (to be chosen by the user) are converted into the byte array, and then its hash value is calculated.

Step 2: We transform the particular hash value into BIG integer number. This number shall be an odd number and is stored in a temp variable. Now this is the number that will basically undergo the given process.

Step 3: Choose two random, very large, prime numbers p & q such that p < temp < q.

Step 4: Choose any of the factor (it could be prime or composite), "f" of Φ (n) = (p – 1) × (q – 1).

It will be very appropriate to choose any of the factors because in the traditional RSA we choose only prime factor and we are providing the flexibility and versatility to choose any of the factor. This will also make it unpredictable even in the easiest case of factorization of RSA.

Step 5: Choose an integer "e," such that (e, f) = 1, In other words we can say that e shall be relatively prime with f.

Step 6: Evaluate the value of integer "d," such that d = e – 1 (mod f). Now we calculate the solution space s: Xf ≡ 1 (mod n)

Step 7: A subset s' of original solution space s in the above step is chosen at random, we can represent the subset s' as f' (0 < f' < f).

And s' is represented in terms of relational set as shown s' = {ai} (ai ε s && 1 <= i <= f). Now we do not need to reveal the value of f because if the value of f (the critical value for cryptanalysis) is not known, then it is impossible to judge the exact value for factorization in the big space of s'.

Step 8: The following keys that will be used are as follows:

- The triple (e, n, s') is considered as public key to encrypt the data.
- The (d, n, s') is considered as private key used for decryption purpose.

Here in the above discussed algorithm it is impractical to get the exact plain text with the help of cipher text. The main reason is that only cipher text and public key is known to the user and to break this, following condition shall satisfy:

A number X such that X Î s', (where ϵ indicates "belong to") so directly searching that particular number from the whole space of s' (i.e., is very large) is very much infeasible.

3.3 Encryption

The sender chooses three main factors of the algorithm:

- The symmetric algorithm to be used for encryption.
- The renewal period of the symmetric keys.

The renewal period or crypto period of the various keys are defined according to NIST standard document (Barker, Baker, Burr, Polk & Smid, 2011) Appendix. In this document it is defined in terms of SUP (Sender usage period) and RUP (Receiver usage period) as reflected in Appendix. In purview of both time periods, largest one is chosen as renewal period or crypto period. The list of various time periods for different time period is given in Appendix.

- The length of the symmetric key.

Now codon sequences are chosen as it will specify where index search is to be done. The negotiation begins when the sender transfers its public key to the corresponding receiver. The public key of receiver is encrypted with the public key of the sender and retransmitted to the sender. The sender will encrypt the chosen parameters with the receiver's public key which he received. The receiver on receiving the parameter might reject them on a factor of providing maximum security. In this case, a renegotiation is done until both are convinced on some security parameters.

The whole encryption process is clearly indicated in the Figure 1. From the programming perspective (used in this research) as shown in the below drawn box, provider public class is providing the basic medium for storing the data in DNA itself. Now various cryptographic techniques have been used to provide the security to this storage class of DNA information. The basic skeleton of programming of encryption and decryption process is as follows:

The basic implementation of a DNA security is as follows:

```
public class DNASecurity extends Provider
{
public DNASecurity
{
super ("DNAProvider", "DNASecurity");
}
}
```

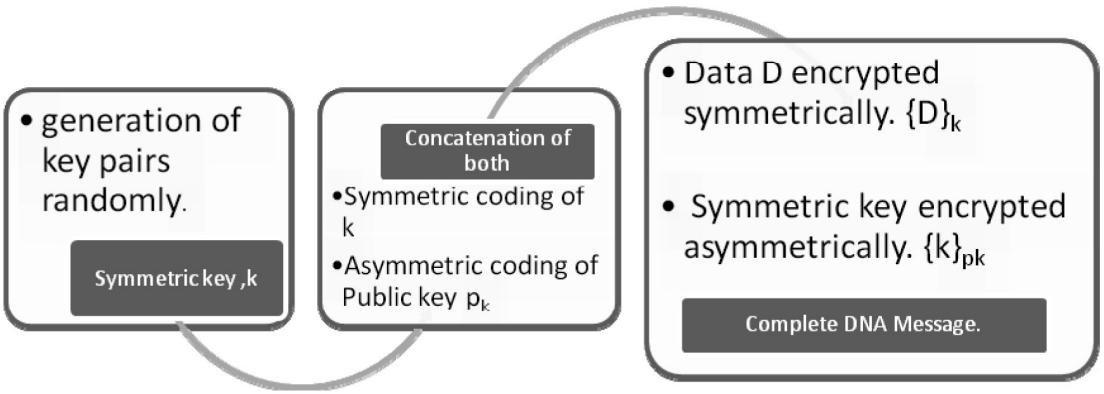

Figure 1 Complete DNA Encryption Process

3.4 Transmission

The encryption phase ends with the transmission of test message, if the transmitted message experiences an error or if too much delay in transmission is present, then the encryption phase needs to be evaluated again. Before the transmission of data can begin, the actual data is encoded using symmetric key, according to the negotiations made earlier. At regular time period "t," symmetric key is generated and receiver's public key is used to encrypt the symmetric key and integrate with original message that has to be transmitted. Henceforth data is encrypted with symmetric key, which is in turn encrypted with receiver's public key. The efficiency and robustness of this approach is even greater than full-fledged algorithm, this is due to the fact that symmetric key is itself encrypted with asymmetric key (through improved version of RSA algorithm)

Now data shall be transformed to a byte array, further it is transformed to raw DNA message. This is done by substitution cipher cryptographic technique by substituting the alphabets. The obtained message is then transformed into indexes i.e., transformed in string form (Figure 1).

3.5 Decryption

The decryption process is much similar to encryption process. The indexes generated in transmission phase are converted into raw DNA array. Now extraction of the data is performed through symmetric key generated by the private key of the receiver and then all negotiations and its data are destroyed.

```
import java.io.*;
import java.lang.String;
    class decryption {
    public static void main (String args []) throws Exception
                                        {
                                            *****Main Body*******
                                        }
                    }
```

Where, D is denoted for data that has to be encrypted, K is denoted as key based on chosen algorithm & P_k is denoted as receiver's public key.

Now according to the Figure 1, complete DNA message is pretty much wrapped with two layers of security. In the first block normally DNA message is encrypted with any of the symmetric key cryptography method then the upper layer security is provided by encrypting the key of symmetric key through asymmetric key cryptography as shown in the second block. Third block shows the complete package of DNA message that has to be transmitted over the insecure network. Thus we are showing in the proposed methodology that we are not only providing the extra layer of security but also providing less complexity (from sender perspective) and high speed (as illustrated in the further section) without compromising the security.

4. Simulation result and analysis

This work clearly indicates, the obtained result stated by previous research work and new work with proposed RSA algorithm (Figure 2) and its comparison with improved version of RSA and other symmetric algorithm (Figure 3). Here we just tested DES, Triple DES, AES, IDEA and improved version of RSA algorithm. The motivation behind this is just to do comparative study of various cryptographic algorithms in terms of time complexity.

Where, time shown on Y axis is measured in milliseconds.

The test message (Figure 4) shown on the console input is used for testing the time complexity. The results shown below (Figure 2 & 3) are checked over stated configuration (Table 1).

In the above stated results (Figure 2 & 3), we are seeing that an exception is being generated i.e., Encryption and Decryption time in DNA cryptography algorithm is on a little bit higher side but we encounter this problem due to java platform used for

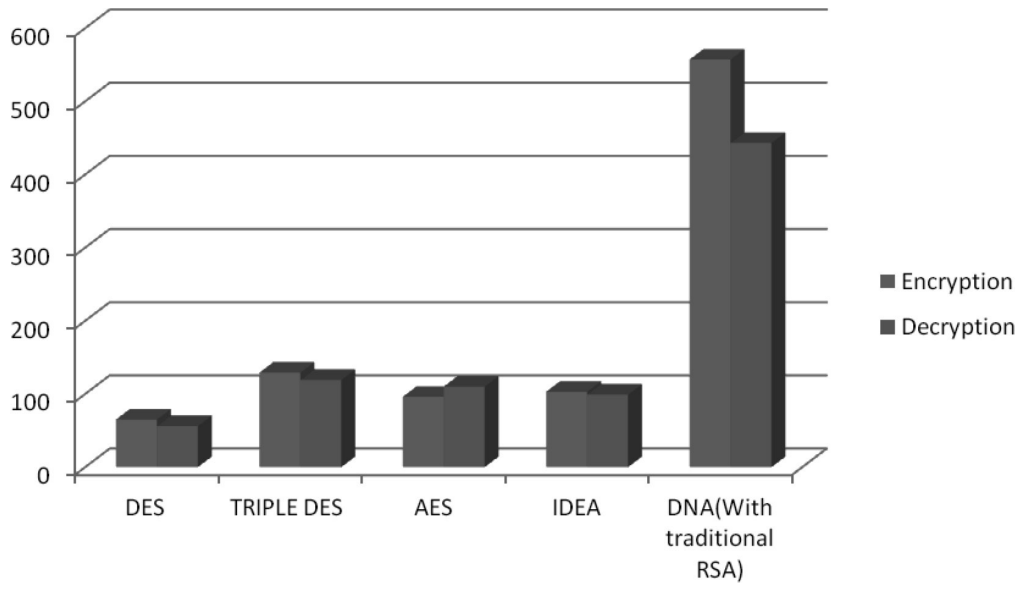

Figure 2 Time Complexity Comparison with Various Cryptographic Algorithms

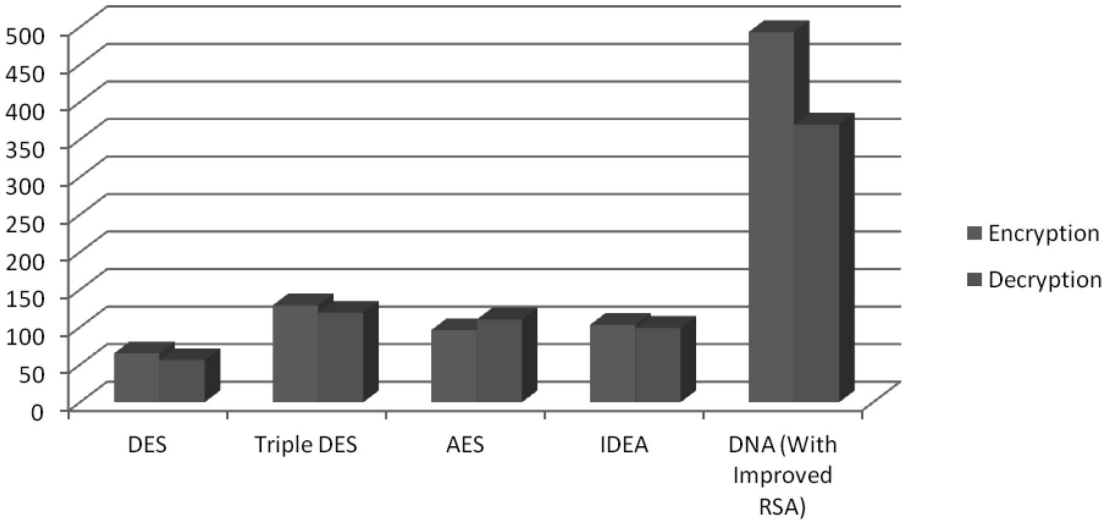

Figure 3 Time Complexity Comparison of Improved RSA (Proposed) with Various Cryptographic Algorithms

simulation. Actually in java, cryptography algorithm is implemented in two parts. First one is public class that could be used by instance of these classes, second is an abstract class; one has to inherit this class that simply needs some extra time. Now, the one more aspect of this simulation result is total computation time that has been calculated through below entered sequences that are shown on the console input (Figure 5).

Table 1 Configuration Used for Testing Time Complexity

Processor	Specification	RAM	Operating System	OS Specification
Intel® Core™ 2 Duo	T5800, 2.00 GHz	1 GB	Windows 8	32 bit

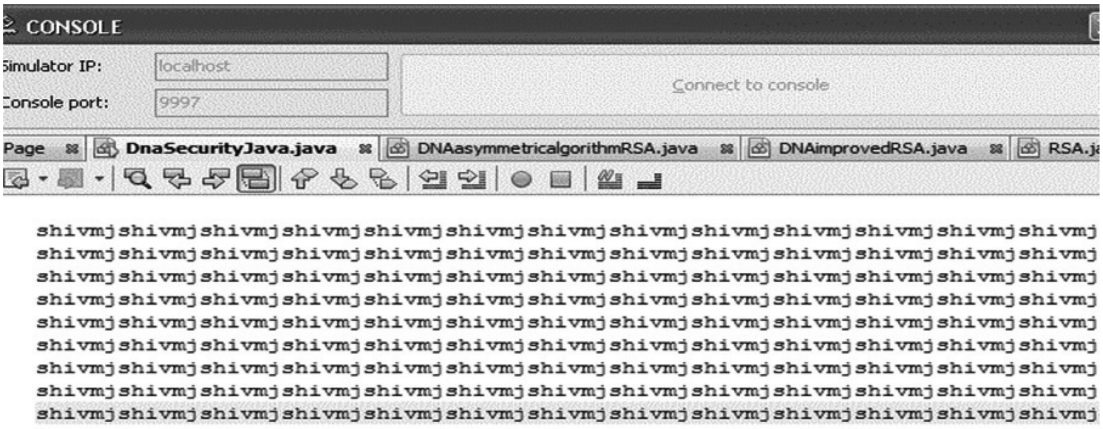

Figure 4 Console Input for Testing Time Complexity

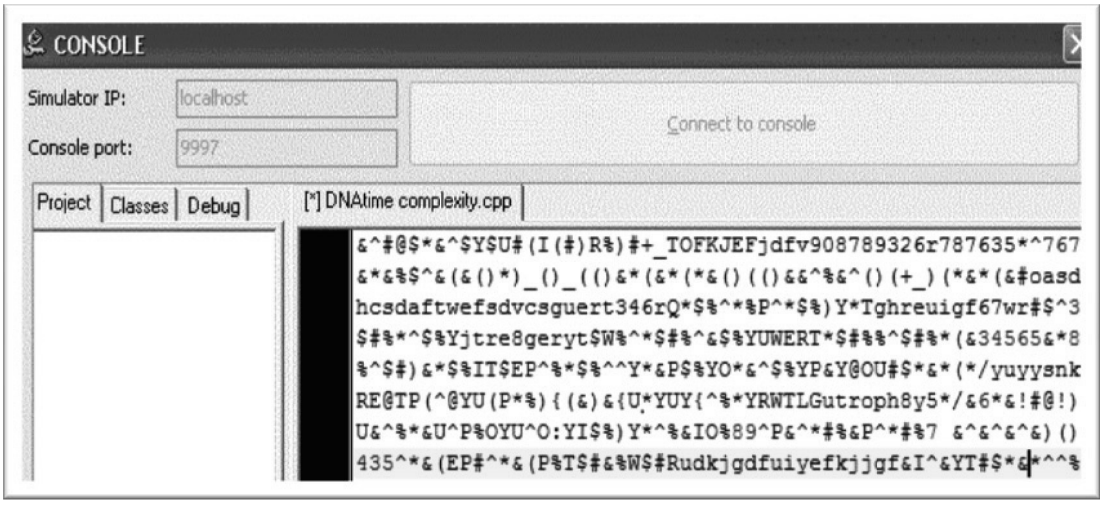

Figure 5 Console Input for Testing Computation Time

It is very important to know for all of us that the total computation time is completely dependent on the OS platform and situation/environment in which we are simulating our results. This could vary according to situation and how this platform is installed. If this is calculated over normal installed platform (Table 2) then we get these results (Table 3).

As shown in the above Table 3, the total computation time of the DNA security algorithm with this hybrid approach (Proposed Algorithm) is very much larger than in

Table 2 Configurations Used for Testing Computation Time

Serial No.	Processor	Specification	RAM	Operating System	Specification
Configuration 1	Intel Dual Core	E5200 @2.50 GHz	2 GB RAM	Windows XP SP 2	32 Bit
Configuration 2	Intel Core i3	M350 @ 2.27 GHz	1 GB RAM	Ubuntu 10.04	Linux

Table 3 Result Obtained for Computation Time for Different Configuration

Algorithm	Configuration 1		Configuration 2	
	Encryption	Decryption	Encryption	Decryption
DES*	1.3	0.77	0.9	0.2
	1.32	0.65	0.62	0.23
	1.36	0.93	1.02	0.46
Triple DES*	4.1	1.12	3.2	0.79
	4.9	1.62	3.9	0.77
	5.2	3.2	3.8	0.72
IDEA*	10	9.3	9.1	5.2
	10.6	7.2	10.4	6.0
	11.2	6.3	11.0	5.8
AES*	1.3	0.13	1.0	0.10
	0.6	0.15	0.5	0.32
	2.3	0.89	2.1	0.93
DNA Security through RSA*	556	6.29	551	2.00
	553	5.2	493	2.62
	439	4.19	425.3	2.32
DNA Security through Improved RSA* (Proposed)	493	4.00	446.7	2.65
	452	3.91	449	3.2
	392.6	3.9	452.1	2.93

Note: *denotes the time taken for encryption and decryption process in "milliseconds."

perfect case of symmetric key cryptography but still less than with traditional asymmetric algorithm (RSA). So this is the whole idea behind DNA information security through public key cryptography consisting of double layers i.e., defense in depth security. DNA encryption and decryption itself is a very complex and time taking process and in this case of implementation (discussed in this paper) object oriented language & JDK platform is used in which abstract classes have been inherited that increase approximately 15% ~ 20% time in encryption and decryption time. Average total computation time in DNA security (proposed algorithm) is 445.866 milliseconds, and this is expected that if this would be

implemented through parallel processing then definitely we shall get computation time reduced.

5. Conclusion and future scope

In this era of information technology, we need a huge and secure space to store our information. As it is discussed in the paper that in a gram of DNA 108 tera byte information is stored and through DNA cryptography we have proposed a very technical, efficient and effective way to secure the information residing in DNA. Although it has been shown that DNA Cryptography is itself a time consuming process and encrypting asymmetrically takes more time(as shown in the results). In the above discussion we proposed a completely new way to encrypt the information contained in DNA by integrating proposed new asymmetric RSA algorithm. This approach provides higher stability (it takes more time to factorize & cryptanalysis) and more powerful (it provides double layered encryption i.e., impractical to break) than any other symmetric algorithm like INDEX based, OTP, DES, and AES. Apart from security concept, it is very clearly demonstrated that windows is more time consuming in terms of computation time in comparison to Linux particularly for the algorithm test. Ultimately in a practical environment we have shown the direct way to manage and secure the information. The problem of excess computation time and less speed can be improved with the help of parallel processing. In future, higher speed can be achieved through parallel processing.

References

Adleman, L.M. (1994), 'Molecular computation of solutions to combinatorial problems', *Science*, Vol. 266, No. 5187, pp. 1021-1024.

Barker, E., Baker, W., Burr, W., Polk, W. and Smid, M. (2012), 'Recommendation for key management -- part 1: general', NIST Special Publication 800-57, National Institute of Standards and Technology, Gaithersburg, MD.

Boneh, D., Dunworth, C. and Lipton, R.J. (1995), 'Breaking DES using a molecular computer', *Proceedings of DIMACS Workshop on DNA Based Computers*, Princeton, NJ, pp. 37-66.

Cherian, A., Raj, S.R. and Abraham, A. (2013), 'A Survey on different DNA cryptographic methods', *International Journal of Science and Research*, Vol. 2, No. 4., pp. 167-169.

Clelland, C.T., Risca, V. and Bancroft, C. (1999), 'Hiding messages in DNA microdots', *Nature*, Vol. 399, pp. 533-534.

Cox, J.P.L. (2001), 'Long-term data storage in DNA', *Trends in Biotechnology*, Vol. 19, No. 7, pp. 247-250.

Cui, G.Z. (2006), 'New direction of data storage: DNA molecular storage technology', *Computer Engineering and Applications*, Vol. 42, pp. 29-32.

Cui, G.Z., Qin, L., Wang, Y. and Zhang, X. (2008), 'An encryption scheme using DNA technology', *Proceedings of IEEE 3rd International Conference on Bio-Inspired Computing: Theories and Applications*, Adelaide, SA, pp. 37-42.

EMBL-EBI. (2012), 'The European Bioinformatics Institute, Part of the European Molecular Biology Laboratory', available at http://www.ebi.ac.uk (accessed 20 December 2013).

Gehani, A., LaBean, T.H. and Reif, J.H. (1999), 'DNA-based cryptography', *Proceedings of 5th Annual DIMACS Meeting on DNA Based Computers*, Cambridge, MA, pp. 233-249.

Lai, X.J., Lu, M.X., Qin, L., Han, J.S. and Fang, X.W. (2010), 'Asymmetric encryption and signature method with DNA technology', *Science China Information Sciences*, Vol. 53, No. 3, pp. 506-514.

Leier, A., Richter, C., Banzhaf, W. and Rauhe, H. (2000), 'Cryptography with DNA binary strands', *BioSystems*, Vol. 57, No. 1, pp. 13-22.

Lu, M.X., Lai, X.J., Xiao, G.Z. and Qin, L. (2007), 'Symmetric-Key cryptosystem with DNA technology', *Science China*, Vol. 50, No. 3, pp. 324-333.

Mills A.P. Jr., Yurke, B. and Platzman P.M. (1999), 'Article for analog vector algebra computation', *BioSystems*, Vol. 52, No. 1-3, pp. 175-180.

Ning, K. (2009), 'A pseudo DNA cryptography method', available at http://arxiv.org/abs/0903.2693 (accessed 25 December 2013).

Nobelis, N., Boudaoud, K. and Riveill, M. (2008), 'Une architecture pour le transfert électronique sécurisé de document', Unpublished PhD dissertation, Equipe Rainbow, Laboratories I3S-CNRS, Sophia-Antipolis, France.

Schena, M. (2003), *Microarray Analysis*, John Wiley & Sons, Hobooken, NJ.

Shimanovsky, B., Feng, J. and Potkonjak, M. (2002), 'Hiding data in DNA', *Proceedings of 5th International Workshop on Information Hiding*, Noordwijkerhout, The Netherlands, pp. 373-386.

Soni, R. and Johar, A. (2012), 'An encryption algorithm for image based on DNA sequence addition operation', *World Journal of Science and Technology*, Vol. 2, No. 3, pp. 67-69.

Tatiana, H., Mircea-Florin, V., Monica, B. and Cosmin, S. (2008), 'A java crypto implementation of DNAProvider featuring complexity in theory and practice', *Proceedings of 30th International Conference on Information Technology Interfaces*, Dubrovnik, Croatia, pp. 607-612.

Terec, R., Vaida, M.F., Alboaie, L. and Chiorean, L. (2011), "DNA security using symmetric and asymmetric cryptography', *International Journal on New Computer Architectures and Their Applications*, Vol. 1, No. 1, pp. 34-51.

Wang, R., Chen, J. and Duan, G. (2011), 'A k-RSA algorithm', *Proceedings of IEEE 3rd International Conference on Communication Software and Networks*, Xi'an, China, pp. 21-24.

Wasiewicz, P., Mulawka, J.J., Rudnichi, W.R. and Lesyng, B. (2000), 'Adding numbers with DNA', *Proceedings of 2000 IEEE International Conference on Systems, Man and Cybernetics*, Nashville, TN, pp. 265-270.

Youssef, I.M., Emam, A. and Abd Elghany, M. (2012), 'Multi-layer data encryption using residue number system in DNA sequence', *International Journal of Security and Its Applications*, Vol. 6, No. 4, pp. 1-12.

About the authors

Shiv P. N. Tripathi is pursuing MS in information Security from Indian Institute of Information Technology, Allahabad (UP) India. In the past, he has completed his B.Tech (Hons.) in Information Technology from BBD National Institute of Technology and Management in 2012. He has worked (as a co-author) in a research paper titled "Basics of an Affordable & Ubiquitous 5G Wireless Network." He has been awarded as an excellent child scientist (as a project leader) in 2006 in National child science congress.

Corresponding author. Division of MS (Cyber Law & Information Security), Indian Institute of Information Technology, Allahabad, Uttar Pradesh, Pin: 211012, India. Tel: +91-9532890146. E-mail address: tripathi161290@gmail.com

Manas Jaiswal is pursuing MS in information Security from Indian Institute of Information Technology, Allahabad (UP) India. In the past, He has completed his B.Tech in Computer Science from Amity University in 2011. He is member of Microsoft Tech. Club sponsored by Microsoft student partner India. E-mail address: manasjswl@gmail.com

Vrijendra Singh is an Associate Professor and a member of academic council in Indian Institute of Information technology, Allahabad (UP), India. He has been a member of various professional and research societies like IEEE, International Association of Engineers. He

has 10+ years of experience in research and teaching field. He has published more than 20 papers in various international or national publications. His major area of interests are Artificial Neural Networks, Information security, Image processing, Cryptography etc. E-mail address: Vrijendra.singh@gmail.com

Appendix

In the NIST document recommended crypto period for specific key types has been given (Barker et al., 2012).

Key Type	SUP	RUP
Private Signature Key		1 ~ 3years
Public Signature Key		Several years (depends on key size)
Symmetric Authentication Key	≤ 2years	≤ SUP + 3years
Private Authentication Key		1 ~ 2years
Public Authentication Key		1 ~ 2years
Symmetric Data Encryption Key	≤ 2years	≤ SUP + 3years
Symmetric Key Wrapping Key	≤ 2years	≤ SUP + 3years
Symmetric and asymmetric RNG Keys		Upon reseeding
Symmetric Master Key		About 1year
Private Key Transport Key		≤ 2years
Public Master Key		1 ~ 2years
Symmetric Key Agreement Key		1 ~ 2years
Private Static Key Agreement Key		1 ~ 2years
Public Static Key Agreement Key		1 ~ 2years
Private Ephemeral Key Agreement Key		One key agreement transaction
Public Ephemeral Key Agreement Key		One key agreement transaction
Symmetric Authorization Key		≤ 2years
Private Authorization Key		≤ 2years
Public Authorization Key		≤ 2years

Source: Barker et al. (2012).

Where, Key denotes the type of key what we are using in cryptographic transaction. SUP denotes the sender's usage period. RUP denotes the receiver's usage period.

MIS Review Vol. 19, No. 1, September (2013), pp. 61-80
DOI: 10.6131/MISR.2013.1901.04

An Effective Pareto Optimality Based Fusion Technique for Information Retrieval

Krishnan Batri
Department of Electronics and Communication Engineering, PSNA College of Engineering and Technology, India

ABSTRACT: *Information Retrieval (IR) is the process of retrieving information that is relevant to the users' needs. Over the years, researchers tend to develop the best retrieval strategy, which achieves the best possible performance across all document collections. Their results indicate a pattern of tug-of-war relationship prevalent among the existing strategies, where in one strategy dominates the remaining strategies over other document collections. Data Fusion may nullify the aforesaid tug-of-war effect. It can extract the best possible performance among the participating members. Data Fusion in IR usually combines the various retrieval schemes (strategies) to enhance the overall system performance. Our proposed fusion functions assign relevance scores by considering non dependency among all participating strategies. Relevance score assignment based on the relationship between that specific document and all other documents in the corpus. The existing Comb functions treated as the baseline functions for our proposed functions. Proposed and baseline functions' performance tested among three medium size corpuses. The average precision value of functions indicates that, one of our proposed functions achieves better performance in comparison with the base line functions. The statistical analysis confirms the same.*

KEYWORDS: *Information Retrieval, Data Fusion, Meta Search, Vector Space Model, Similarity Measures, Extended Boolean Model.*

1. Introduction

Knowledge sharing helps the mankind in the evolution process. Once they start sharing their knowledge, they became more civilized. Knowledge sharing process made up of two critical tasks. The first one deals with the creation and the second one associated with the sharing. Apart from knowledge creation, knowledge sharing seems to be more critical as it indirectly preserves the knowledge. Hence, this process rendering an important help for the welfare of mankind. Information Retrieval (IR) is the process of sharing the knowledge among the needy people. We may call it as science, art, or a technique, but it meticulously store, organize, and proffer the required information.

Information retrieval is the process of retrieving the required information according to the users' needs (Baeze-Yates & Ribeiro-Neto, 1999; Korfhage, 1997; Salton & McGill, 1983). According to Yates, "Information retrieval deals with the representation, storage, organization of and access to the information items (Baeze-Yates & Ribeiro-Neto). The representation and organization of the information items should provide the user with easy access to the information in which he or she is interested."

The information retrieval process made up of three important tasks. The first one deals with the representation of the information. As the available information may be in structured, semi structured and unstructured format, they should be represented in a common format. In order to carry out this task, some preprocessing mechanisms have to be carried out. As the pre-processing mechanisms being the foundation step, variations in their performance alter the overall system's performance. Hence, researchers focused more towards the pre-processing mechanisms and contribute more. Plenty of works carried out on some of the pre-processing mechanisms like, stemming, tokenization, and stop word removal. Based on their results, we came to a final conclusion that the performances of these pre-processing mechanisms are not unique. They tend to vary from one application to other.

Organizing and storing of keywords lies at the middle level. Information should be organized in a manner that the matching process becomes an easier one. This task termed as the indexing process. The indexing process should discriminate the keywords, and store them in a proper manner. The discrimination process involves with the assignment of weights to all keywords. Hence, the indexing process involves with the storing of index terms along with their calculated weights.

The critical nature of the indexing process proves to be more vital than the preprocessing step. Hence, plenty of works carried out in the indexing mechanisms. These works critically divided in to two main categories. The first one deals with the data structure, which used to store the index terms. The second one involves with the calculation of index terms' weights. The weighting mechanism is more critical as it has the ability to alter the overall system's performance. It should discriminate a keyword from the rest. Method of calculating the index term's weight vary from one approach to other. Hence, the performance of the weighting mechanism is not unique. It varies from one application to other.

Method of calculating the correlation between the user's query and the information source is the last task in information retrieval process. Actually the retrieval system finds the relevant information in this last task because the information source available prior to the query, and they are indexed. The query posted at the last minute. The previous two tasks assist the last one. In other words, these three tasks are dependent. They should be

executed in a sequential order. If there is a deviation in any one of the task, it will affect the entire retrieval process.

The corpus may contain one or more relevant information sources. If it has only one source, than there won't be any problem. If there are few or more, than one important question will surface out. Which one is more relevant? Answer to this question demands a measuring method, which used to measure the correlation between the query, and the information sources. It termed as the similarity measure. The principle of operation of a similarity measure depends on the underlying storage model and the weight assignment mechanism. Hence, the literature flooded with different types of similarity measures.

Once the retrieval systems calculate the degree of correlation, the next question will pop up. How to list the relevant sources? Enumeration of the relevant sources should be based on their degree of correlation. Hence, the final task subdivided in to two sub tasks. The first sub task calculates the correlation and the second one list the sources based on their correlation. In some retrieval system, there is no provision for listing. Apart from this, the different retrieval systems use the different weighting schemes. As the performance of the weighting schemes is not consistent, the performance of the retrieval system also varies.

Different methods used to implement the three different retrieval tasks of the retrieval process. As the choices are plenty, there is a need for standardization. Various models used to standardize the retrieval process. The retrieval model expresses the method of processing, organizing, storing, and retrieving the information. Based on their operating principle, the models classified in to three types. They are (1) Exact match model, (2) Vector space model, and (3) Language model. Each of these models has its own advantages, and disadvantages. These models incorporate the three primary tasks of the retrieval processes differently. Hence, the performance of these retrieval models not consistent.

From the above discussions, we come to a conclusion that, performance of retrieval models and the three tasks is not consistent. It is varying. If we propose a better model or mechanism near future, it will also render an inconsistent performance. Even more, our proposed mechanism will lose the battle ground against some other new models or mechanisms. In engineering, there is a scope for improvement. Hence, instead of spending our energy to develop a new model, why can't we tap the positive potential of these existing models and mechanisms. Answer to this leads to the development of new research area called data fusion. The data fusion merges the merits of underlying mechanisms or models. If a better model or mechanism evolves, we can add it to the pool. Hence the data fusion has some scope of enhancement. Hence, the data fusion seems to be better than the individual models or schemes. It proves to be better.

Rest of this article organized as follows. Section 2 gives the details about the retrieval models and data fusion principles. Section 3 gives the insight about the earlier works in the area of information retrieval and data fusion. Section 4 gives the details about our proposed work. Section 5 gives the details about the experimental setup and the results. Section 6 concludes with the future direction of our research.

2. Models of IR and data fusion

This section dedicated to IR models and data fusion. Various types of models and their underlying principles, their comparison are discussed in the first part of this section. The concept of data fusion along with its needs, and its principles are given in the last part of this section.

2.1 IR models

IR models used to describe the principles associated with each, and sub tasks of the retrieval process. More specifically a *model* is a set of premises and an algorithm for ranking documents with regard to a user query (Salton & McGill, 1983). More formally, an IR model is a quadruple [D, Q, F, $R(q_i, d_j)$] where D and Q is a set of logical views of documents and queries and $R(q_i, d_j)$ is a ranking function which associates a numeric ranking to the query q_i and the document d_j and F is the frame work for modelling document and queries. *Strategy* or scheme is synonymous with rank $R(q_i, d_j)$. It is a method of assigning similarity between the query and the documents. *System* refers to the physical implementation of an IR algorithm which can have various operational modes or various settings of parameters. Therefore the same IR system may be used to execute different IR schemes by adjusting the various parameters.

Performance of the IR system depends on the underlying IR algorithm, which in turn depends on the underlying IR model. IR models classified in to three types. Out of which, exact match model is most primitive. Lots of works carried out on the vector space model. It is almost saturated. Language models are still in developing state. Hence, works are going on in the language model. Hence, we focus more towards the first two models. In future, we plan to accommodate the language model.

2.1.1 Exact match model

Boolean model is very simple, and it operates on the principle of Boolean algebra. It retrieves documents based on a word matching function. Since the decision space in Boolean Model is binary, documents judged as either relevant or irrelevant. Thus the number of documents retrieved as a result of the Boolean nature of the model is either vast or too small. Also there is no provision for the ranking the documents. These limitations

eliminated by extending the Boolean Model with the functionality of partial matching and term weighting. This extended model combines the advantages of the Boolean model and the VSM.

Salton introduced the Extended Boolean Model (EBM) on 1983. In this model, the weights assigned to the terms lie between zero to one. It uses the maximum normalization method, and the normalized weights assigned to the index terms. The function used in maximum normalization given in Equation (1).

$$normalized\ w_{i,j} = \frac{unnormalized\ w_{i,j}}{w_l} \tag{1}$$

Where,

$w_{i,j}$ = weight of the term i in j^{th} document and

w_l = maximum weight of the generic index term l in the corpus.

The weight assignment techniques in the EBM are same as that of VSM with the only difference being that the weights are normalized. The matching function or similarity measure adapted from the Boolean Model. In Extended Boolean Model, a query represented in one of the following forms: (1) Conjunctive form, (2) Disjunctive form, and (3) Combination of both conjunctive and disjunctive form.

In disjunctive query form, distance from (0,0) used as the similarity measure between the query and the document. Conjunctive query form uses (1,1) as the origin for the distance measure. The distance measure not restricted to Euclidean distance but generalized to any value ranging from 1 to ∞. As EBM depends on the value of p (distance) for calculating the similarity value, it also referred as P-norm model. The generalized form of the query in conjunctive and disjunctive form is represented in Equations (2) and (3) respectively.

$$q_{or} = (w_1 \vee^p, w_2 \vee^p, w_3 \vee^p, \cdots\cdots\cdots, w_m \vee^p) \tag{2}$$

$$q_{and} = (w_1 \wedge^p, w_2 \wedge^p, w_3 \wedge^p, \cdots\cdots\cdots, w_m \wedge^p) \tag{3}$$

The similarity measure between the document, and the query in the P-norm model given in Equations (4) and (5).

$$Sim(q_{or}, d_j) = \left(\frac{w_1^p + w_2^p + w_3^p + \cdots + w_m^p}{m}\right)^{1/p} \tag{4}$$

$$Sim(q_{and}, d_j) = 1 - \left(\frac{(1-w_1)^p + (1-w_2)^p + \cdots + (1-w_m)^p}{m}\right)^{1/p} \tag{5}$$

Where,

w_m = weight of the index term and $1 \leq p \leq \infty$

2.1.2 Vector Space Model

Vector Space Model (VSM) is the most popular IR model. VSM not only explains the process of retrieving relevant documents but also the assignment of rank to the documents. In VSM, the objects of IR, such as term, document, and query treated as multidimensional linearly dependent vectors in the vector space.

In vector space model, the weight $w_{t,d}$ associated with the index term "t" in a document "d" is positive and non binary. Furthermore, the index terms in the query also weighted. Let $w_{t,q}$ be the weight associated with the pair [t,q], where $w_{t,q} \geq 0$. Then the query vector q is defined in Equation (6).

$$q = (w_{1,q}, w_{2,q}, w_{3,q}, w_{l,q}, \ldots, w_{n,q}) \tag{6}$$

Where,

n = the total number of index terms in the system,

$w_{l,q}$ = weight of the index term l in query q.

In VSM, the documents represented as a linear combination of keywords or index terms. The weights of the index terms can be calculated in many ways. Since the objects of IR treated as being linearly dependent, term vector can be represented as a linear combination of documents. The term vector defined as follows:

$$t_l = (d_1, d_2, d_3, \ldots, d_N), \quad 1 \leq l \leq n$$

Where,

d_N = weight of the t_l^{th} term in the d_N^{th} document,

N = total number of documents in the corpus and

n = total number of index terms in the corpus.

The scalar product between the query, and document vectors used to calculate the relevance (similarity) of the document with respect to the query. In the vector space V, if there exist two vectors x and y such that $x, y \in V$ then the scalar product defined in Equation (7).

$$S(x, y) = |x| \cdot |y| \cdot \cos\theta \tag{7}$$

Where,

|x|, |y| = magnitude of the vectors,

$| x | = \sqrt{\sum_{i=1}^{n} x_i^2}$,

$| y | = \sqrt{\sum_{i=1}^{n} y_i^2}$ and,

θ = angle between two vectors.

The vector space that considers only the scalar product termed as the Euclidean space. The scalar product used as one of the methods to calculate the correlation between the query, and the document vectors. There are various methods available to calculate the correlation (similarity) value. Based on the value of the correlation between the query and the document, the relevance of the document justified. The retrieved documents arranged in descending order based on the value of their similarity.

2.2 Data fusion

Effectiveness of the existing IR system depends on the underlying model and strategy. Certain strategies perform well in a specific environment while their performance deteriorates in other environments. Early research shows that there is no single strategy that achieves constant performance across all test document collections (Zobel & Moffat, 1988). As a result of the ever increasing users of Internet, and hence the massive information repository, there is a challenging research requirement to work out a strategy whose effectiveness should be high compared to the other existing retrieval strategies.

Recent research work identifies that fusion technique improves and stabilize the IR system performance (Fox & Shaw, 1994; 1995; Lee, 1995; 1997a; 1997b). Fusion is the methodology of combining retrieval strategies associated with the retrieval task followed by an assignment of relevance score or rank to documents on the basis of the score returned by the fused strategies (Bartell, Cottrell & Belew, 1994; Belkin, Kantor, Cool & Quatrain, 1994; Vogt, 1999). Fusion methods broadly classified into: (1) Data fusion, and (2) Collection fusion. The detailed classification of fusion techniques is shown in Figure 1.

The data fusion approaches combine the results obtained from various retrieval strategies over the same document collection or corpus, whereas collection fusion combines the results of various document collections. In Collection fusion, the same query operates over the various document collections. The relevant documents returned from the multiple corpuses merged together to give the final relevant documents list. In IR, same document, and query can be represented using different weighting scheme. If the

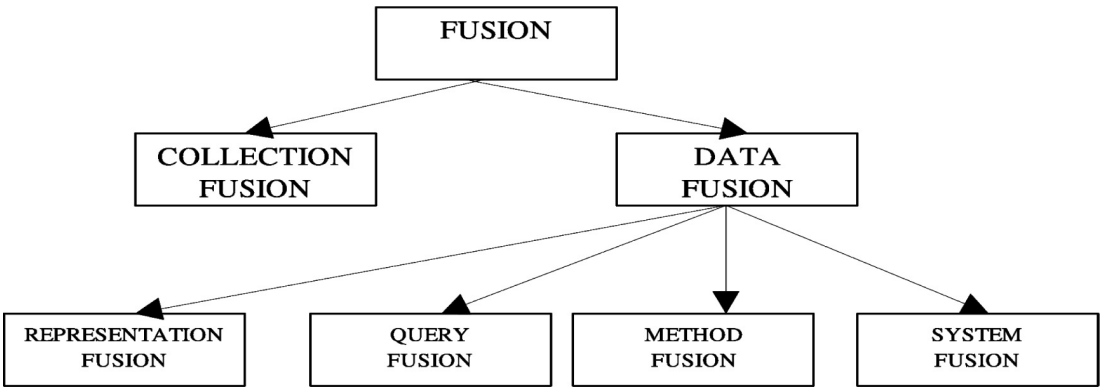

Figure 1 Types of Fusion Techniques

fusion operation merges the result of the various document representations then it termed as representation fusion. If the various query forms fused then it termed as query fusion. Various methods used to retrieve relevant document, and if these methods fused together then it termed as method fusion. The results from the multiple systems merged then it is termed as system fusion. Previous results show that, the fusion methods render some positive impact over the effectiveness of the retrieval system. It also yields consistent results over all test document collections.

3. Prior work in fusion technique

This section details on the various research studies that have been carried out on *Meta Search Algorithm for Combining Scores*. Researchers have exploited the characteristics of "*Meta Search Algorithm for Combining Scores*" by fusing the various existing IR models. Meta Search Algorithm combines the relevance scores returned from the various retrieval strategies to identify the relevant documents.

Early work on data fusion method that does not use training data commenced in the early 70's when Fisher (Fisher & Elchesen, 1972) fused two Boolean searches together. In his method, one search operates on the title word while the other search operates on manually generated index terms. He achieved significant improvement in effectiveness of the information retrieval system. This method combines minimum number of retrieval strategies whereas linear combination method successfully combines more number of strategies. The linear combination method assigns weights to the individual strategies (Vogt, 2000; Vogt & Cottrell, 1999). The final relevance score of a document assigned by weighted linear combination method is given in Equation (8).

$$R(q, d) = \sum_{i=1}^{k} \theta_i \cdot E_i(q, d) \qquad (8)$$

Where,

θ_i = Weight of the i^{th} retrieval strategy,

E_i (q,d) = Relevance score returned by the i^{th} retrieval strategy and

k = Number of retrieval strategies to be fused.

The weighted linear combination method has the limitation of requiring prior knowledge about the retrieval systems to assign the weights. This limitation is eliminated in Comb-functions by treating all strategies equally.

The Comb functions for combining scores by treating all strategies equally have been proposed by Fox and Shaw (Fox & Shaw, 1994). The various Comb Functions used for combining scores is shown in Figure 2.

Fusion techniques differ from Comb-functions in the fact that the relevance is computed on the basis of rank assigned to documents as compared to the relevance scores methodology adopted in Comb-functions. Few such fusion techniques that emulate the social voting schemes are Boarda fusion and Condorcet fusion (Montague & Aslam, 2002). Extensive work on Comb functions has been carried out by Lee. New rationales and indicators for data fusion have been proposed by Lee. He has conducted experiments over TREC data collection. He concluded that CombMNZ is the better performing function than the other Comb-functions.

The training data involved in the fusion techniques are used to assign weights to individual strategies. The weighted scores from the individual strategies are combined linearly to assign the final relevance score. The weighted linear combination method maintains the same weight for all retrieval systems. But the performance of the system differs from query to query. Hence the selection of the best performing retrieval strategy becomes vital. Probabilistic approach is used for this purpose. The best performing strategy is selected automatically from the pool. The probabilistic model selects only one strategy from the pool and all other strategies become idle. Hence evolutionary algorithms are used to select the best performing strategies (Coello, 2000). Billhardt, Borrajo and Maojo (2003) proposed a heuristic based data fusion algorithm. They uses Genetic

CombMIN	*Minimum of Individual Similarities*
CombMAX	*Maximum of Individual Similarities*
CombSUM	*Summation of Individual Similarities*
CombANZ	*CombSUM ÷ Number of Nonzero Similarities*
CombMNZ	*CombSUM × Number of Nonzero Similarities*

Figure 2 Comb-functions for Combining Scores

algorithm to combine the retrieval score. Their algorithm not only assigns the scores to independent strategies but also selects the best performing strategy for fusion.

Fusion techniques utilize the advantages of its member strategies by combining the strategies together. By combining the strategies it tends to exploit the following effects as indicated by Vogt (1999): (1) Skimming effect, (2) Chorus effect and (3) Dark Horse effect. The *Skimming Effect* happens when retrieval approaches that represent their collection items differently may retrieve different relevant items, so that a combination method that takes the top ranked items from each of the retrieval approaches will push non-relevant items down in the ranking. The *Chorus Effect* occurs when several retrieval approaches suggest that an item is relevant to a query; this tends to be a stronger evidence for relevance than that of a single approach. The *Dark Horse Effect* is one in which a retrieval approach may produce unusually accurate (or inaccurate) estimates of relevance for at least some items, relative to the other retrieval approaches. By carefully designing the combination function, researchers utilized the advantages of these above said effects. Our proposed function employs the advantages of skimming effect.

The performance of the fusion techniques that requires training data depends on the relevance feedback from the active participant. So the performance of such systems differs from user to user and depends on their skill of predicting relevance of the documents. Hence we concentrate on complete user independent fusion techniques. Literature survey has established Comb-functions to be the best performing functions in this category. In this work, we compare the performance of our proposed functions over the CombMNZ function, the best among the Comb-functions.

4. Proposed work

This section discusses about our *Pareto Optimality* based fusion technique and its comparison with (1) the existing CombMNZ, a best Meta search algorithm for combining scores and (2) remaining Comb functions detailed in the previous section. A Meta search algorithm combines the results obtained from more than one data source. In IR, fusion algorithm operates on various retrieval strategies to calculate the final relevance score of the document. These retrieval strategies act as the criteria for selection of relevant document. The Decision Vector for Multi Criteria Selection in IR is represented mathematically as in Equation (9).

$$s_i(q, d) = \{s_i^1(q, d), s_i^2(q, d), \ldots\ldots, s_i^j(q, d)\} \tag{9}$$

Where,

i = document index

j = number of retrieval strategies to be fused

s_i^j (q, d) = relevance score returned by the j^{th} retrieval strategy.

In multi criteria selection, similarity values of a document from various retrieval strategies are treated as an independent variable of a decision vector. Hence the final decision about the relevance of the document cannot be arrived at by operating only on the vector space. Relevance of the document is decided by calculating the equivalent scalar value of the decision vector. We use the notion of *Pareto Optimality* to calculate the equivalent scalar value (FRS). According to Pareto optimality, in a maximization problem, a vector $s_i^* \in V$ is said to be Pareto optimal "if all other vectors have smaller value for at least one retrieval strategy or have the same value for all retrieval strategies." In other words

s_i^* is said to be Pareto Optimal, iff $\forall j \; s_i^j = s_k^j$ *or*

at least one value of l such that $l \in j, \; s_i^j > s_k^l$

There are various methods available to calculate the final scalar value. Our proposed approach treats all retrieval strategies equally. In order to maintain equality we normalize the relevance scores and use maximum normalization function. The mathematical equation to calculate the normalized score under maximum normalization is represented in Equation (10).

$$s_{normalized} = \frac{s_{unnormalized}}{s_{\max}} \tag{10}$$

Where,

$s_{unnormalized}$ = relevance score returned by a retrieval strategy and

s_{max} = maximum relevance score returned by a generic retrieval strategy.

In our proposed method of combining relevance scores, assignment of final relevance score to a document is based on the relationship between the corresponding document and all other remaining documents in the corpus. We choose the difference between the scores with respect to each of the document for the same retrieval strategy as a metric to establish a relationship. The relevance score difference between documents obtained via all retrieval strategies are of two types namely: (1) Minimum and (2) Maximum difference. The relationship between two documents based on the above mentioned relevance score difference is expressed mathematically as in Equations (11) and (12).

$$d_i \cong X + d_j \quad j = 1, 2, 3, \ldots\ldots N,\ j \neq i \tag{11}$$

$$d_i \cong Y + d_j \quad j = 1, 2, 3, \ldots\ldots N,\ j \neq i \tag{12}$$

Where,

d_i = a specific document to be compared with the other remaining document,

X = minimum difference between two documents in all retrieval strategies,

Y = maximum difference between two documents in all retrieval strategies and

N = total number of documents in the corpus.

For an example, take the relevance score returned by the four retrieval strategies for document 1 and 2 as

d_1 = {5,4,3,2}

d_2 = {1,2,3,1}

Now calculate the difference between the document 1 and 2

$d_1 - d_2$ = {4,2,0,1}

Now take the maximum and minimum value from the above calculated difference

Maximum difference (Y) = 4

Minimum difference (X) = 0

Now the relationship between the document 1 and 2 is

$d_1 \cong 4 + d_2$

$d_1 = 0 + d_2$

Based on the value of minimum and maximum difference, we establish the relationship for a specific document with all other remaining documents in the corpus. The relationship space R of a specific document consists of (N-1) minimum and (N-1) maximum differences. From the relationship space R, we find out either the maximum or minimum value that is globally optimal. We choose one of the global minimum or maximum. The mathematical representation of global maximum and global minimum is given in Equations (13) and (14).

$$x_i \prec x_j \quad iff \quad \forall j(x_i < x_j), i \neq j,\ j = 1, 2, \ldots (N-1). \tag{13}$$

$$x_i \prec x_j \quad iff \quad \forall j(x_i > x_j), i \neq j,\ j = 1, 2, \ldots (N-1). \tag{14}$$

Based on the local and global relationships, we derive four functions to assign the final relevance score to the documents. The proposed functions are (1) C-maxmax, (2) C-maxmin, (3) C-minmax and (4) C-minmin. The formulas used to assign final relevance score based on C functions are given in Figure 3.

One of the main advantages of the proposed method is that it does not require any weight assignment and training data. Since our proposed approach treats all strategies equally, our proposed approach exploits the advantages of "*skimming effect*."

5. Experimental results

This section details on the experimental results of our proposed functions which were discussed in the previous section. We conducted the experiments on three test document collections, namely, (1) CRANFIELD, (2) CISI and (3) ADI under an uniform environment. The document abstracts in CRAN collection are about Aeronautics and these documents are compiled by cranfield institute of technology. The CISI dataset is about library science and it is collected by Institute of Scientific Information. ADI is a small data set in the field of Information Science. Table 1 shows the characteristics of the three

$$FRS_{C-\max\max} = \max_{\forall j, j\neq k}\left(\max_{i=1,2..n}\left(s_i^k - s_i^j\right)\right)$$

$$FRS_{C-\max\min} = \max_{\forall j, j\neq k}\left(\min_{i=1,2..n}\left(s_i^k - s_i^j\right)\right)$$

$$FRS_{C-\min\max} = \min_{\forall j, j\neq k}\left(\max_{i=1,2..n}\left(s_i^k - s_i^j\right)\right)$$

$$FRS_{C-\min\min} = \min_{\forall j, j\neq k}\left(\min_{i=1,2..n}\left(s_i^k - s_i^j\right)\right)$$

Where,
j,k = Document id,
i = retrieval strategy
s_i^j = relevance score of j^{th} document in i^{th} retrieval strategy

Figure 3 C-functions for Combining Scores

Table 1 Characteristics of Datasets

Characteristics	ADI	CISI	MED
Number of documents	82	1,460	1,033
Number of terms	374	5,743	5,831
Number of queries	35	35	30
Average number of document relevant to a query	5	8	23
Average number of terms per document	45	56	50
Average number of terms per query	5	8	10

datasets. We measured the 11 point interpolated precision to judge the performance of our retrieval strategy.

5.1 Retrieval Strategies

Retrieval Strategy is used to assign similarity score between the document and the query. We use various similarity measures of VSM and P-norm model as retrieval strategies and fuse them. Various similarity measures of VSM used in our experiment are given in Equations (15) ~ (18).

$$\text{Cosine Similarity } R(q,d) = \frac{\sum_{i \in q \cap d} w_{q,t} \cdot w_{d,t}}{W_q \cdot W_d} \tag{15}$$

$$\text{Inner Product } R(q,d) = \sum_{i \in q \cap d} w_{q,t} \cdot w_{d,t} \tag{16}$$

$$\text{Dice Coefficient } R(q,d) = \frac{\sum_{i \in q \cap d} w_{q,t} \cdot w_{d,t}}{W_q^2 + W_d^2} \tag{17}$$

$$\text{Jaccard } R(q,d) = \frac{\sum_{i \in q \cap d} w_{q,t} \cdot w_{d,t}}{W_q^2 + W_d^2 - \sum_{i \in q \cap d} w_{q,t} \cdot w_{d,t}} \tag{18}$$

Where,

R: Relevance score of document d with respect to query q,

$w_{q,d}$: weight of the term t in the query q,

$w_{d,t}$: weight of the term t in the document d,

W_q: weight of the query and

W_d: weight of the document d.

The conjunctive query form of P-norm model mentioned in Equation (5) is also used as a retrieval strategy in our experiment. We use p value as 1.5, 2.5 and 3.5 to calculate the similarity score in P-norm model. We use the above seven retrieval strategies to test the effectiveness of our proposed functions over (1) CombMNZ, the best meta-search algorithm used for fusion and (2) remaining Comb functions.

We maintain a uniform environment by using the same stop-word list, stemmer algorithm and weight assignment mechanism. The formulas used to assign weight to index terms are given in Equations (19) ~ (21). We use Term Frequency-Inverse Document Frequency (TF-IDF) weight assignment schemes for assigning weights to index terms.

$$w_t = \log_{10}\left(1+\frac{N}{f_t}\right) \tag{19}$$

$$w_{d,t} = r_{d,t} \cdot w_t \tag{20}$$

$$r_{d,t} = f_{d,t} \tag{21}$$

Where,

w_t = term weight

$w_{d,t}$ = document term weight

$r_{d,t}$ = relative term frequency

$f_{d,t}$ = frequency of the term t in document d

5.2 Results

We conducted the experiments by combining the various similarity measures of VSM along with the P-norm similarity measures. We chose P value as 1.5, 2.5, and 3.5. We used 11 point interpolated precision to calculate the effectiveness of the proposed functions. We also used the average value of 11-point interpolated precision to compare the effectiveness of our proposed functions against the Comb functions. The results for the proposed C-functions and the Comb functions are shown in Figure 4.

The averages of 11 point interpolated precision are given in Table 2 to make the comparison process easy. The last column shows the average of all functions over all document collections. The table indicates that C-minmax is the best performing function compared to remaining C functions and Comb functions; it achieves 5.95% improvement over the CombMNZ function.

The CombMNZ function is the subset of linear combination model. The linear combination model assigns weights to the retrieval strategies exploiting the *Chorus effect*. CombMNZ function treats all strategies equally and assigns equal weights to all

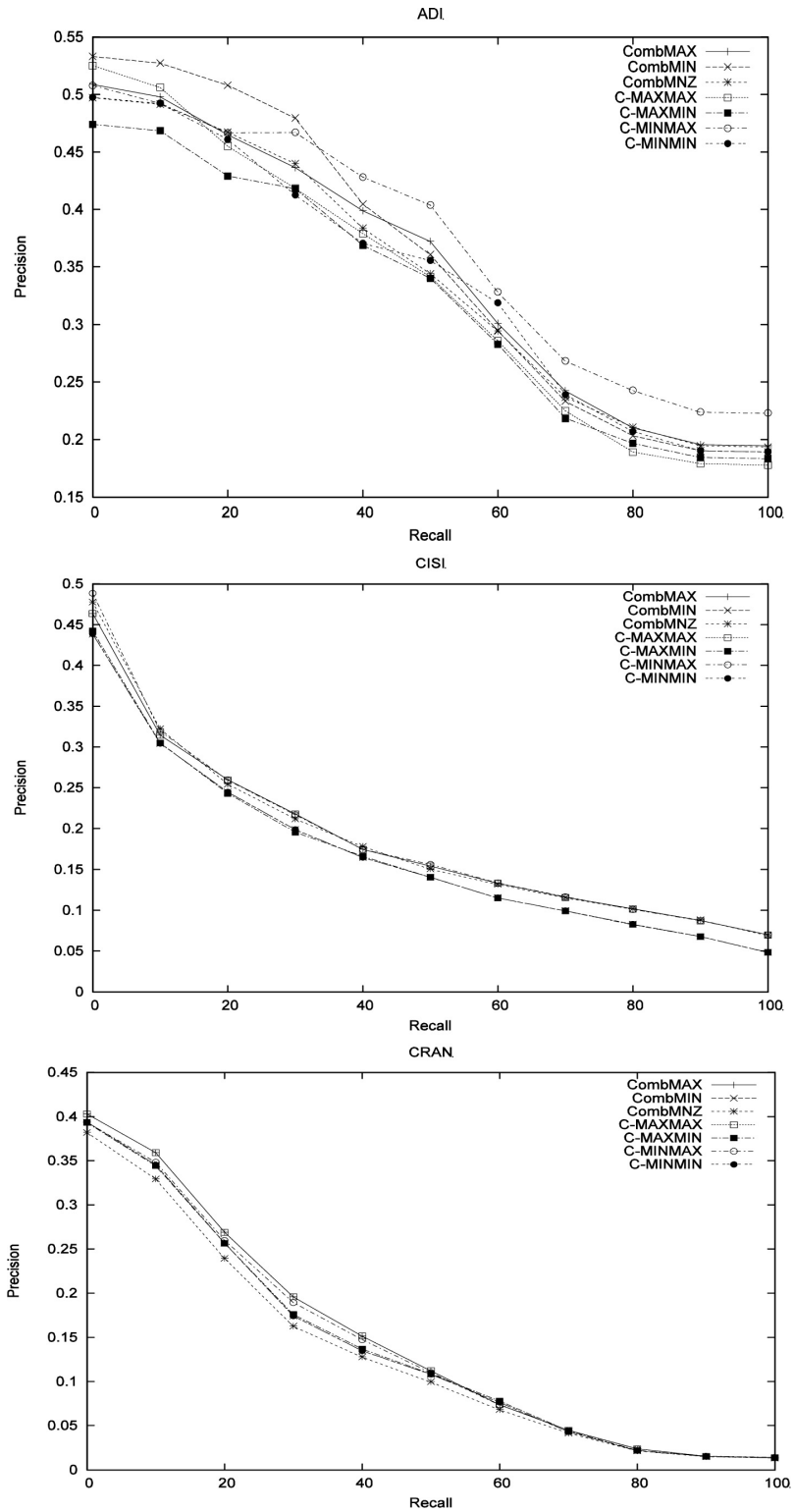

Figure 4 11-pt Interpolated Precision Curve for C and Comb Functions

Table 2 Average 11-pt Interpolated Precision for C and Comb Functions

Function	ADI	CISI	CRAN	Average
CombMNZ	0.3413	0.1911	0.1364	0.2229
CombMAX	0.3475	0.1901	0.1511	0.2296
CombMIN	0.3569	0.1733	0.1439	0.2247
C-Maxmax	0.3347	0.1901	0.1511	0.2253
C-Maxmin	0.3240	0.1732	0.1444	0.2139
C-Minmax	0.3685	0.1929	0.1473	0.2362
C-Minmin	0.3395	0.1733	0.1439	0.2189

retrieval strategies. Since CombMNZ assigns equal weights to all strategies and it is one of the subsets of linear combination model, it exploits the advantages of both chorus and skimming effects. Our proposed C-functions incorporate skimming effect, since it treats all strategies equally. The proposed approach does not combine the scores linearly and the final relevance score depends on the individual scores as compared to the existing linear combination of scores methodology adopted in CombMNZ. Also the proposed approach is based only on skimming effect whereas CombMNZ utilizes both skimming and chorus effects. The limitation of such a strategy is that a large Chorus effect cuts into the possible gain from the skimming effect, thereby leading to degradation in performance.

In our proposed functions of C-maxmax and C-minmax, the right half that is $\max_{i=1,2..n}\left(s_i^k - s_i^j\right)$ part maximizes the difference between the similarity scores returned from the corresponding retrieval strategies. The ideal relevance score of a relevant document is set as one and for the non relevant it is set as zero. Hence the maximum allowed difference value is one. Since the sub portion $\max_{i=1,2..n}\left(s_i^k - s_i^j\right)$ maximizes the difference, it indirectly chooses the document which has the relevance score more close to the ideal value. As a result both of the C-maxmax and C-minmax functions perform well in our experiment. But the slight degradation in performance of C-maxmax function is due to the Dark horse effect. Few retrieval strategies unexpectedly give maximum scores to non relevant documents. C-maxmax chooses the maximum value from the calculated difference as compared to C-minmax which chooses the minimum difference. Hence C-minmax becomes the best performing function compared to the remaining C-functions and Comb functions.

6. Conclusion

We have proposed a set of new functions for combining multiple relevance scores in information retrieval. The proposed functions do not require any training data and return only the relevance scores as compared to ranks being returned by other existing fusion methods. The average value of the 11 point interpolated precision over the three test document collections shows that the C-minmax is the better performing function compared to remaining C-functions and Comb functions. The proposed functions treat all strategies equally like that of Combfunctions. The proposed approach compute the relationship between the documents, hence it require some extra computation time. The C-functions introduced in this paper is very much useful for medium size document collection.

References

Baeze-Yates, R. and Ribeiro-Neto, B. (1999), *Modern Information Retrieval*, Addison Wesley, Harlow, UK.

Bartell, B.T., Cottrell, G.W. and Belew, R.K. (1994), 'Automatic combination of multiple ranked retrieval systems', *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Dublin, Ireland, July.

Belkin, N.J., Kantor, P., Cool, C. and Quatrain, R. (1994), 'Combining evidence for information retrieval', *Proceedings of the Second Text REtieval Conference*, Gaithersburg, MD, August/September.

Billhardt, H., Borrajo, D. and Maojo, V. (2003), 'Learning retrieval expert combinations with genetic algorithm', *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 11, No. 1, pp. 87-113.

Coello, C.A. (2000), 'An updated survey of GA-based multiobjective optimization techniques', *ACM Computing Survey*, Vol. 32, No. 2, pp. 109-143.

Fisher, H.L. and Elchesen, D.R. (1972), 'Effectiveness of combining title words and index terms in machine retrieval searches', *Nature*, Vol. 238, pp. 109-110.

Fox, E.A. and Shaw, J.A. (1994), 'Combination of multiple searches', *Proceedings of the Second Text REtieval Conference*, Gaithersburg, MD, August/September.

Fox, E.A. and Shaw, J.A. (1995), 'Combination of multiple searches', *Proceedings of the Third Text REtieval Conference*, Gaithersburg, MD, November.

Korfhage, R.R. (1997), *Information Storage and Retrieval*, John Wiley & Sons, New York, NY.

Lee, J.H. (1995), 'Combining multiple evidence from different properties of weighting schemes', *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Seattle, WA, July.

Lee, J.H. (1997a), 'Analyses of multiple evidence combination', *Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Philadelphia, PA, July.

Lee, J.H. (1997b), 'Combining multiple evidence from different relevant feedback networks', *Proceedings of the 5th International Conference on Database Systems for Advanced Applications*, Melbourne, Australia, April.

Montague, M.H. and Aslam, J.A. (2002), 'Condorcet fusion for improved retrieval', *Proceedings of the 11th International Conference on Information and Knowledge Management*, McLean, VA, November.

Salton, G. and McGill, M.J. (1983), *Introduction to Modern Information Retrieval*, McGraw-Gill, New York, NY.

Vogt, C.C. (1999), 'Adaptive combination of evidence for information retrieval', Unpublished doctoral dissertation, University of California, San Diego, CA.

Vogt, C.C. (2000), 'How much more is better? Characterizing the effects of adding more IR systems to a combination', *Proceedings of RIAO 2000: 6th International Conference on Content-Assisted Information Retrieval*, Paris, France, April.

Vogt, C.C. and Cottrell, G.W. (1999), 'Fusion via a linear combination of scores', *Information Retrieval*, Vol. 3, No. 1, pp. 151-173.

Zobel, J. and Moffat, A. (1988), 'Exploring the similarity space', *ACM SIGIR Forum*, Vol. 1, No. 32, pp. 18-34.

About the author

Krishnan Batri is Professor at Department of Electronics and Communication Engineering, PSNA College of Engineering and Technology, Dindigul. He has completed hid Ph.D from National Institute of Technlogy, Trichy on 2008. His area of interest includes Information Retrieval, Text Mining, and Genetic Algorithm. His research papers published in various international journal and conferences.

Corresponding author. Department of Electronics and Communication Engineering, PSNA College of Engineering and Technology, Muthanampatty, Dindigul, TamilNadu, India. Tel: +91-9789680969. E-mail address: krishnan.batri@gmail.com

CALL FOR PAPER

MIS Review: An International Journal

Published 2 Issues Annually by Airiti Press Inc.

MIS Review is a double-blind refereed academic journal published jointly by Airiti Press Inc. and Department of Management Information Systems, College of Commerce, National Chengchi University in Taiwan. The journal is published both in print and online. We welcome submissions of research papers/case studies in the areas including (but not limited to):

1. MIS Roles, Trends, and Research Methods

Roles, positioning and research methods of management information systems, and the impacts & development trends of information technology on organizations.

2. Information Management

Information infrastructure planning and implementation, information technology and organizational design, strategic applications of information systems, information system project management, knowledge management, electronic commerce, end-user computing, and service technology management.

3. Information Technologies

Database design and management, decision support systems, artificial intelligence applications (including expert systems and neural networks), software engineering, distribution systems, communication networks, multimedia systems, man-machine interface, knowledge acquisition & management, data mining, data warehouse, cooperative technology, and service science & engineering.

4. Information Applications and Innovations

The applications and innovations of business functional information systems (e.g., production, marketing, financial, human resources, and accounting information systems), enterprise resource planning, customer relationship management, supply chain management, intellectual capital, geographic information systems, and integrated information systems.

5. Information Technology Education and Society

Information education, e-learning, and information impacts on society.

6. Others

Other MIS-related topics.

INSTRUCTIONS FOR SUBMISSION

1. Papers can be prepared in either Chinese or English. If your paper is written in Chinese, it will be translated into English once it is accepted for publication.
2. There is no submission deadline for MIS Review. All papers will be double-blind reviewed by at least two reviewers, who will be recommended by the Editorial Board. The processing time for the first-round formal reviews is about six weeks. Subsequently rounds of reviews tend to be faster.
3. To simplify file conversion effort, PDF or Microsoft Word 2000/2003 (for Windows) format is advised. Then, please submit your paper via the MIS Review website (URL: http://www.icebnet.org/misr/).
4. MIS Review is an academic journal. According to international practice, once an article is accepted and published, MIS Review will not give or take any payment for the publishing. An electronic copy of the paper will be sent to the article author(s) for non-profit usage.
5. The submitted and accepted paper should follow the author guidelines for paper submission format provided on the MIS Review website.

The submitted paper should include the title page, abstract, key words, the paper body, references, and/or appendices. You must submit three files. The information of author(s) should not appear anywhere in the paper body file, including page header and footer.

1. On a separate (cover letter) file, please follow the author guidelines provided on the MIS Review website to prepare the letter.
2. On a separate (title page) file, please note the title of the paper, names of authors, affiliations, addresses, phone numbers, fax numbers, and E-mail addresses.
3. On a separate (paper body) file, please include the paper title, an abstract, a list of keywords, the paper body, the references, and/or appendices. The abstract must contain the research questions, purposes, research methods, and research findings. The abstract should not exceed 500 words and the number of keywords must be 5-10 words.
4. The submitted and accepted paper should follow the author guidelines for paper submission format provided on the MIS Review website.

CONTACT

Editorial Assistant
Department of Management Information Systems
College of Commerce, National Chengchi University
No. 64, Sec. 2, ZhiNan Road, Wenshan District,
Taipei 11605, Taiwan R.O.C.
Phone: +886-2-29393091 ext.89055
E-mail: misr@mis.nccu.edu.tw

airiti press Subscription Form

MIS Review

You may subscribe to the journals by completing this form and sending it by fax or e-mail to

Address: 18F., No. 80, Sec. 1, Chenggong Rd., Yonghe District, New Taipei City 23452, Taiwan (R.O.C.)
Tel: +886-2-29266006 ext. 8301 Fax: +886-2-29235151 E-mail: press@airiti.com Website: http://www.airitipress.com

PERSONAL				LIBRARIES / INSTITUTIONS			
	Europe	US/CA	Asian/Pacific		Europe	US/CA	Asian/Pacific
1 Issue	€ 34	US$ 41	US$ 38	1 Issue	€ 53	US$ 65	US$ 62

Vol.		No.		~	Vol.		No.		Copies		US$		Total US$	

*All Price include postage

PLEASE NOTE

- Issues will be sent in two business days after receiving your payment.
- Please note that all orders must be confirmed by fax or email.
- Prices and proposed publication dates are subject to change without notice.
- Institutions include libraries, government offices, businesses, and for individuals where the company pays for the subscription.
- Personal rates are available only to single-user personal subscribers for personal and non-commercial purposes.
- Airiti Press reserves its right to take appropriate action to recover any losses arising fromany intended or unintended misrepresentation of the term "Personal Subscriber".

BILLING INFORMATION

Name	
Company	
Tel	Fax
E-mail	
Shipping Address	

INTERNATIONAL PAYMENTS

Pay by Credit Card	
Card Type ☐JCB ☐MasterCard ☐Visa	
Card Name	
Card Number	
Expiry Date _____ / _____	CVV number
Signature	

Direct Bank Transfer	
Beneficiary	AIRITI INC.
Address	18F., No. 80, Sec. 1, Chenggong Rd., Yonghe District, New Taipei City 23452, Taiwan (R.O.C.)
Bank Name	E.Sun Commercial Bank, Ltd.Yong He Branch
Account No	0107441863017
Swift Code	ESUNTWTP
Bank Address	No.145, Zhongzheng Rd., Yonghe District, New Taipei City 23454, Taiwan (R.O.C.)