THE IIOAB JOURNAL

VOLUME 9 : NO 3 : OCTOBER 2018 : ISSN 0976-3104

Institute of Integrative Omics and Applied Biotechnology Journal

Dear Esteemed Readers, Authors, and Colleagues,

I hope this letter finds you in good health and high spirits. It is my distinct pleasure to address you as the Editor-in-Chief of Integrative Omics and Applied Biotechnology (IIOAB) Journal, a multidisciplinary scientific journal that has always placed a profound emphasis on nurturing the involvement of young scientists and championing the significance of an interdisciplinary approach.

At Integrative Omics and Applied Biotechnology (IIOAB) Journal, we firmly believe in the transformative power of science and innovation, and we recognize that it is the vigor and enthusiasm of young minds that often drive the most groundbreaking discoveries. We actively encourage students, early-career researchers, and scientists to submit their work and engage in meaningful discourse within the pages of our journal. We take pride in providing a platform for these emerging researchers to share their novel ideas and findings with the broader scientific community.

In today's rapidly evolving scientific landscape, it is increasingly evident that the challenges we face require a collaborative and interdisciplinary approach. The most complex problems demand a diverse set of perspectives and expertise. Integrative Omics and Applied Biotechnology (IIOAB) Journal has consistently promoted and celebrated this multidisciplinary ethos. We believe that by crossing traditional disciplinary boundaries, we can unlock new avenues for discovery, innovation, and progress. This philosophy has been at the heart of our journal's mission, and we remain dedicated to publishing research that exemplifies the power of interdisciplinary collaboration.

Our journal continues to serve as a hub for knowledge exchange, providing a platform for researchers from various fields to come together and share their insights, experiences, and research outcomes. The collaborative spirit within our community is truly inspiring, and I am immensely proud of the role that IIOAB journal plays in fostering such partnerships.

As we move forward, I encourage each and every one of you to continue supporting our mission. Whether you are a seasoned researcher, a young scientist embarking on your career, or a reader with a thirst for knowledge, your involvement in our journal is invaluable. By working together and embracing interdisciplinary perspectives, we can address the most pressing challenges facing humanity, from climate change and public health to technological advancements and social issues.

I would like to extend my gratitude to our authors, reviewers, editorial board members, and readers for their unwavering support. Your dedication is what makes IIOAB Journal the thriving scientific community it is today. Together, we will continue to explore the frontiers of knowledge and pioneer new approaches to solving the world's most complex problems.

Thank you for being a part of our journey, and for your commitment to advancing science through the pages of IIOAB Journal.

Yours sincerely,

Vasco Azevedo

Vasco Azevedo, Editor-in-Chief
Integrative Omics and Applied Biotechnology
(IIOAB) Journal

Prof. Vasco Azevedo
Federal University of Minas Gerais
Brazil

Editor-in-Chief

Integrative Omics and Applied Biotechnology (IIOAB) Journal Editorial Board:

ARTICLE

CBPLSA - AN EFFECTIVE COLLABORATIVE FILTERING ALGORITHM FOR DISTRIBUTED DATA MINING ON ELECTRONIC HEALTH RECORDS

S. Urmela[1*], M. Nandhini[2]

[1]*Department of Computer Science, Pondicherry University, Puducherry, INDIA*

ISSN 0976-3104

ABSTRACT

Background: *In recent years, Distributed Data Mining (DDM) on Electronic Health Records (EHRs) has become one of dominant area in DM.* ***Methods:*** *This paper proposed architecture for EHRs DDM using memory-based and model-based collaborative filtering. The proposed CBPLSA(Cluster-Based Probabilistic Latent Semantic Analysis) algorithm for DDM on EHRs which aims to minimize computational complexity and memory overhead by maintaining clusters of old patients' EHRs.* ***Implementation:*** *Experimental implementation on real-world EHRs datasets available on Hypertension, Diabetes and Meningitis depicts an improved precision and result accuracy compared to state-of-arts EHRs retrieval approaches.*

KEY WORDS
Distributed Data Mining
Electronic Health Records
CBPLSA algorithm
Collaborative Filtering
Clustering

Received: 18 Feb 2018
Accepted: 31 Mar 2018
Published: 8 April 2018

***Corresponding Author**
Email:
urmelaindra@gmail.com
Tel.: +91 9791958868

INTRODUCTION

Data mining (DM) is the process of extracting useful, important information using patterns from available datasets.[1] discusses that the tremendous growth of information and technology has paved way to explore other forms of data mining which includes Collective Data Mining (CDM/DDM), temporal data mining, phenomenal data mining and visual data mining. DDM is devoted to retrieve patterns from distributed datasets. Centralized data mining performs data computation at a dedicated geographical location. Processing data at centralized site paves way for questioning on privacy of sensitive data, increased computation cost, memory cost and transmission cost. The objective of DDM is to extract useful, unknown information from heterogeneous data sites. It involves computation at heterogeneous points, hosting individual computing units. De-centralized data mining also makes entire system scalable, by distributing the workload among heterogeneous computing points. Traditional algorithms developed for DM mostly devoted for centralized environments has proven to be unsuitable for DDM [1]. Alfredo Cuzzocrea(2013)[2] states that framing a methodology for DDM is challenging not only by distributed environment, but also for its efficient resource sharing and minimizing computational complexity specifications. Kargupta et al. [3] and Zaki M J et al.[4] discussed that several researchers analyzed complexity involved in framing methodology for DDM in two ways:

- Analyzing on effective and efficient usage of computational resources at individual distributed data-sites.
- Performing knowledge discovery at individual distributed data site (local level).
- Aggregating knowledge discovered at global level.

Byung-Hoon Park et al. [5] developed an architecture for DDM where processing takes place locally at individual data sites. Finally data will be accumulated to form global model. Grigorios Tsoumakas et al. [6] presented an architecture for DDM where knowledge acquired from local distributed data sites are accumulated at global level forming a merger site. Fu Y et al.[7] discussed certain issues in developing DDM algorithms namely formulating suitable DDM algorithm for heterogeneity datasets; minimizing computational complexity and space complexity; communication cost; privacy preservation of data at distributed site; data fragmentation; data replication and maintaining local datasets autonomy. Further all these issues are interrelated to each other. This has given pave to many researchers to carry-out their work in this field mainly on EHRs retrieval.

This work is inspiring us to bring out a high-performance contribution towards DDM by framing a suitable retrieval algorithm to more interesting but partially explored areas like health informatics, e-science and bioinformatics. Mainly in recent years, in the field of health informatics, retrieval of EHRs with DDM is targeting about minimizing computational cost and computational complexity. Research works so-far proposed on EHRs retrieval have covered various retrieval approaches by clustering, classification and association [2]. Inspired by this, to minimize the computational and memory costs in retrieval, we propose an algorithm for EHRs retrieval approach with memory-based CF and model-based CF (CBPLSA algorithm) for DDM.

RELATED WORKS

A number of prominent EHR mining in centralized environment has been designed. But the performance of mining in centralized environment alone is too limited and paves way for universal EHR. By implementing universal EHR, increased health care cost in terms of repeated laboratory tests can be avoided, promotes effective clinical-decision making.

IIOAB JOURNAL
ISSN: 0976-3104

DDM on EHR

Some of the distributed mining approaches on EHR are discussed below.

Mohammed Khalilia et al.[8] discussed a disease prediction framework based on random forest (RF), an ensemble homogeneous classifier approach based on repeated random sampling of trained datasets. National Inpatient Sample (NIS) data obtained through Healthcare Cost and Utilization Project (HCUP) was used for experimentation. Eight chronic diseases were predicted and performance was compared with other ensemble classifier approach, namely bagging and boosting, along with VM (Vector Machine).

Yan Li et al.[9] discussed a novel security-based distributed ensemble classifier approach for predicting model for EHR data. Each participating homogeneous sites will accumulate dataset in local level. Finally, at global level prediction model will be generated from multiple local models.

CF on health prediction

Some of the CF approaches on health prediction are discussed below.

Martin Lopez et al.[10] discussed a new CF, Property-based Collaborative Filtering (PBCF), introduced links among semantic properties which handles both users and items. This approach helped to solve CF problems namely, scalability, by enabling developer to build a matrix of how much a feature for one person influences the same feature item for some another person. PBCF was used in health-aware recommender system. It was implemented in HARE (Heath-Aware Recommender)[10], a system introduced to deliver personalized ads. PBCF was implemented in HARE, as a health-aware recommender system.

Davis Darcy A et al.[11] proposed Collaborative Assessment and Recommendation Engine (CARE) which uses patients' medical history and similarity among patients' for predicting future greater disease risks. Also an iterative version, ICARE was devised which uses ensemble, homogeneous classifier approaches for achieving better performance.

Kai Zheng et al.[12] discussed Electronic Medical Record Search Engine (EMERSE), implemented along with collaborative search by preserving knowledge of collected EHR search and circulated among other EHR searches. This approach involves use of social-gathered information, helping in accurate and efficient health care information retrieval. Complex search terms and smaller EHR search engine size are certain cons. Above discussed works of DDM on EHRs, focus mainly on privacy of EHRs data and effective clinical decision support system. The proposed work on EHRs retrieval by memory and model-based CF (CBPLSA algorithm – Cluster-Based Probabilistic Latent Semantic Analysis) for DDM aims to minimize computational complexity and to increase EHRs result accuracy. In the next section, problem definition of proposed EHRs retrieval approach is discussed.

PROBLEM DEFINITION

The proposed work of this paper involves framing EHR retrieval, with model-based and memory-based Collaborative Filtering. Among three model-based CF approaches discussed by Koren Y[13], clustering CF along with Probabilistic Latent Semantic Analysis is implemented. As Xiaoyuan Su et al. [14] discussed on hybrid CF, meta-level hybridization technique model learnt from model-based CF on EHRs will be applied over memory- based CF EHRs. Memory-based CF involves EHRs retrieval by framing cluster by CBPLSA algorithm of old patients' EHRs. In model-based CF (new patients EHRs retrieval), clustering model of EHRs which is applied over memory-based CF is formulated. For each disease to be queried corresponding latent words (say for hypertension datasets, latent words are SBP and DBP) will be identified and its probability values will be fixed by analyzing corresponding datasets. Corresponding values of latent words with higher probability will be considered for checking of inclusion of EHRs. The proposed work will be implemented on three publicly available, real-world datasets obtained from Department of Biostatistics, Vanderbilt University. [Table 1] depicts datasets considered, number of patient records and variables along with number of distributed sites considered.

Table 1: EHR Datasets

Datasets	Hypertension	Diabetes	Meningitis
No. of records	381	403	310
No. of variables	05	19	43
No. of distributed sites	08	02	05

PLSA on EHR

Probability of latent words for each disease to be queried will be calculated using the given formulae[15],

$$P(Di \mid Q) = P(Di) * P(Q \mid Di)$$

where, $P(Di) = 1.0$

$$P(Q \mid Di) = P(Q \cap Di) / P(Di)$$

IIOAB JOURNAL
ISSN: 0976-3104

$$P(Q \cap Di) = \sum_{i=1}^{n} P(Li \mid Di) * P(Q \mid Li)$$

where, $P(Q \mid Li) = P(Q \cap Li) / P(Li)$

$P(Li) = 1.0$

In above formulae P(Di | Q) denotes probability of occurrence of keyword Q in corresponding dataset Di of disease queried which is related to probability of dataset considered P(Di) and probability of occurrence of keyword in corresponding datasets P(Q | Di). P(Q | Di) is formulated by calculation of probability of corresponding latent words of disease queried P(Q | Li) and probability of occurrence of latent words in corresponding disease datasets P(Li | Di) (here i, denotes corresponding latent words for disease queried). Here P(Q ∩ Li) represents calculating probability of latent word corresponding to user query. If identified latent word corresponds to Q then P(Q ∩ Li) is 1 else 0. P(Li | Di) represents calculating occurrence of latent word in dataset. If corresponding latent word occurs in dataset then P(Li | Di) is 1 else 0.

ARCHITECTURAL OVERVIEW

The proposed architectural model on EHR mining involves three-phase operations namely, i) EHR retrieval by memory-based CF ii) EHR retrieval by model-based CF iii) meta-level hybridization technique on retrieved EHR.

In phase I, memory-based CF involves forming cluster of distinct disease identified from history of EHRs by CBPLSA algorithm discussed in algorithm 1. In phase II, retrieval of EHRs by model-based CF (CBPLSA algorithm) on new patients' is done. In phase III, meta-level hybridization technique as stated by Xiaoyuan Su et al.[14] involves model learned from model-based CF (CBPLSA algorithm) being applied on memory-based CF. Further it employs exclusion of redundant EHRs by removing matched patient ID on EHRs retrieved in phase I and phase II.

CBPLSA Algorithm

CBPLSA algorithm is designed with PLSA followed by cluster formation of EHR[15].

Algorithm 1. CBPLSA algorithm

Input: query Q, Latent words Li, Dataset considered Di
Output: Set of EHR Clusters

// Retrieve EHR by PLSA
Initialize i = 1;
For each Li **do**
Initialize P(Di) = 1.0;
Initialize P(Li) = 1.0;
Calculate P(Di | Q) = P(Di) * P(Q | Di);
Compute P(Q | Di) = P(Q ∩ Di) / P(Di);
Compute P(Q | Li) = P(Q ∩ Li) / P(Li);
Compute P(Q ∩ Di) = P(Li | Di) * P(Q | Li);
i++;

// Cluster formation of EHR
Intialize f = 1;
For each Lf **do**
Compute corresponding values of Lf for each Q;
Formulate Cluster Cf;
f++;

Algorithm 1 of CBPLSA algorithm reveals EHR retrieval by PLSA along with cluster formation. At initial level, for diseases queried, corresponding latent words with higher probability values for each query (disease type queried namely diabetes, hypertension and meningitis) will be identified by PLSA. In next level, those EHR records' with higher or lower values (defined by medical experts) of latent words is formed as cluster (cluster formed within certain range). After that with next latent words values, subsequent cluster will be formed. For disease to be queried, higher probability latent words of corresponding disease have been considered. Based on medical experts' opinion on the abnormal range of corresponding values of latent words (eg: for hypertension datasets latent words are SBP and DBP with normal range of SBP <140 and DBP<90), clusters of EHRs are formulated. For understanding, in-case of hypertension datasets, cluster C1 is EHRs with SBP and DBP in range >290/200, C2 in range <290/200 to >240/150, C3 in range <240/150 to >190/100 and C4 in range <190/100 to >140/90.

EHR retrieval by memory-based CF

In phase I for old patients' record retrieval, memory-based CF is applied. It involves 3-stages namely,

IIOAB JOURNAL
ISSN: 0976-3104

cluster formation by CBPLSA algorithm, matrix formulation and query matching.

Cluster formation by CBPLSA

Clusters for each disease queried have to be formed by CBPLSA algorithm. Latent word analysis, CBPLSA algorithm and cluster formation are the modules for memory-based CF using CBPLSA algorithm.

Matrix formulation

Matrix is formulated [15] with EHRs query Q(example hypertension, diabetes and meningitis) and cluster of EHR records' and is exhibited in [Fig. 1]. A √, tick mark indicates inclusion of corresponding cluster for respective EHRs query and a X, cross mark indicates exclusion of corresponding cluster for respective EHRs query.

	C1	C2	C3	C4	C5
Q1	X	X	√	X	X
Q2	X	√	X	X	X
Q3	X	X	X	√	X
Q4	√	X	X	X	X
Q5	X	X	X	X	√

Fig. 1: Matrix formulation with EHRs query and cluster

Query matching

By memory-based CF, old patients' EHR records' are retrieved by analysing user query Q with formulated matrix. Corresponding clusters computed by CBPLSA algorithm has been retrieved. From [Fig. 1] if query is Q1 (diabetes) among 5 clusters C3 cluster falls under corresponding disease category, similarly for query Q2 (meningitis) among 5 clusters C2 falls under corresponding disease category. Clusters are matched EHRs records'.

EHR retrieval by model-based CF

In phase II, for new patients' record retrieval, model-based CF is applied. CBPLSA algorithm was applied like on-the-fly over new retrieved EHRs. Same phases as EHR retrieval by memory-based CF have to be applied except the difference that clusters formulation to be done using CBPLSA algorithm only in model-based CF.

The major difference between memory-based CF and model-based CF is inclusion of old patients' EHRs for memory-based CF (EHRs retrieval by cluster formation by latent word analysis, matrix formulation and query matching) and new patients' EHRs for model-based CF (EHRs retrieval by latent word analysis and cluster formulation). Memory-based CF involves prediction/retrieval from history/previous preferences/records whereas model-based CF involves prediction/retrieval by on-the-fly approach.

Removing redundant EHRs

In phase III, EHRs retrieved by memory-based CF and model-based CF are accumulated and any redundant EHRs will be excluded for matched patient-ID from both CFs as defined [15]. The final distinctive result of EHRs is corresponding to user query is obtained. To prove its efficiency, this proposed architecture is implemented and their performance is analyzed with state-of-art datasets [Fig. 2].

EXPERIMENTAL IMPLEMENTATION

In this section, implementation detail is presented, followed by description of comparison works and probability values of latent words of EHRs is depicted.

Experimental setup

Experiment was implemented with the proposed architecture using real-world EHRs obtained from Department of Biostatistics, Vanderbilt University. EHR datasets include Hypertension, diabetes and meningitis patient EHRs. Experiment is implemented using C# language and more effective interface is designed which displays latent words along with probability values for disease queried. The results are obtained by implementing three state-of-arts approach of EHR retrieval along with the proposed architecture.

ISSN: 0976-3104

Comparison works

Proposed architecture using CBPLSA algorithm for DDM on EHRs is compared with 3 state-of-arts EHR retrieval approach namely, DDM approach on EHR, memory-based (old patients EHRs) CF on EHR and model-based (new patients EHRs) CF on EHR. Experimental results were analyzed on all the three EHRs datasets considered.

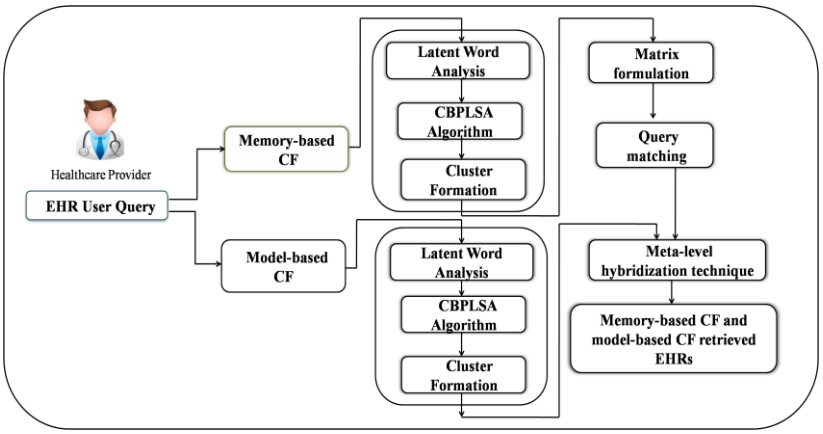


Fig. 2: Architecture - Memory-based and model-based CF (CBPLSA algorithm) for DDM on EHRs retrieval

Performance evaluation

Performance metrics are used for determining the executing capability of proposed program on datasets considered along with computation time required. Evaluation of experimental observations on proposed architecture is analyzed on effectiveness and efficiency measures.

Effectiveness measures

(i) Precision(P): Precision value which is calculated as follows represents a measure that EHRs retrieved is relevant to medical provider query[15].

$$\text{Precision (P)} = \frac{TP}{TP + FP}$$

(ii) Recall(R): Recall value which is calculated as follows represents that relevant EHRs is retrieved by medical provider query[15].

$$\text{Recall (R)} = \frac{TP}{TP + FN}$$

(iii) F-measure: F-measure value which is calculated as follows represents that a higher value of F-measure indicates higher precision and recall values[15].

$$\text{F-measure} = 2 * \frac{Precision * Recall}{Precision + Recall}$$

(iv) Result Accuracy: Accuracy value which is calculated as follows indicates performance of algorithm based on medical provider query[15].

$$\text{Result Accuracy} = \frac{TP + TN}{\text{Total Records}}$$

Efficiency measures

(i) Execution Time: Execution time of proposed algorithm and state-of-arts EHRs retrieval approach are compared by varying the number of EHRs.
(ii) Scalability: Scalability involves calculation of computation time in varying the EHRs datasets. Scalability and execution time goes hand in hand.
(iii) Memory Cost: Memory cost involves computation of memory consumed in running proposed algorithm and state-of-arts EHRs retrieval approach by varying the number of EHRs.

Experiment is implemented with proposed architecture, latent words probability values along with effectiveness and efficiency measures are calculated for 3 EHRs datasets.

Calculation of probability values

Latent words for each disease are fixed based on information obtained from medical experts. Probability values are fixed by PLSA. [Table 2] depicts latent words along with its probability values for each disease. By assigning latent words probability values for each disease, experiment is simulated and performance is analyzed on all the 3 state-of-art approaches along with proposed approach.

IIOAB JOURNAL
ISSN: 0976-3104

Table 2: Latent words and its probability values

Disease	Latent word	Probability
Hypertension	SBP (Systolic Blood Pressure)	0.5389
Hypertension	DBP (diastolic Blood Pressure)	0.6756
Diabetes	chol (Total Cholesterol Level)	0.7645
Diabetes	glyhb (Glycosolated hemoglobin)	0.8956
Diabetes	HDL (High Density Lipoprotein)	0.7845
Meningitis	whites (Total leukocytes)	0.7854
Meningitis	Polys	0.9863

Effectiveness measures observations

[Table 3] depicts the calculated effectiveness measures for the proposed approach along with the comparison works. From [Table 3], the combined approach of memory and model-based CF using CBPLSA algorithm for DDM on EHRs shows higher value of precision, recall and F-measure compared to memory based and model-based CF on EHRs retrieval approaches. The keyword based approach of DDM on EHR shows higher precision, recall and F-measure. As entire dataset is searched sequentially for EHRs retrieval in this case, leads to maximum computational complexity whereas with the clusters formation from history of EHRs, the proposed EHRs retrieval architecture by memory-based CF and model-based CF for DDM has minimized computational complexity and memory overhead.

Table 3: Experimental results. A - keyword based DDM on EHR, B - DDM + memory-based CF on EHR, C - DDM + model-based CF on EHR, D – Proposed CF using CBPLSA for DDM on HER

EHRs Query	Metrics (%)	A	B	C	D
Hyper tension	Precision	1	0.83	0.73	1
	Recall	0.94	0.79	0.54	0.87
	F-measure	0.96	0.80	0.62	0.93
Diabetes	Precision	1	0.91	0.69	1
	Recall	0.95	0.64	0.57	0.93
	F-measure	0.97	0.75	0.62	0.96
Meningitis	Precision	1	0.88	0.82	1
	Recall	0.94	0.61	0.65	0.79
	F-measure	0.96	0.72	0.72	0.88

In [Fig. 3] result accuracy of all the 3 datasets considered along with proposed architecture in-comparison with 3 state-of-arts EHRs retrieval approach is depicted. Approach A has result accuracy value of 85%, approach B of 68%, approach C of 77% and proposed EHRs retrieval by CBPLSA algorithm of DDM of 91%.

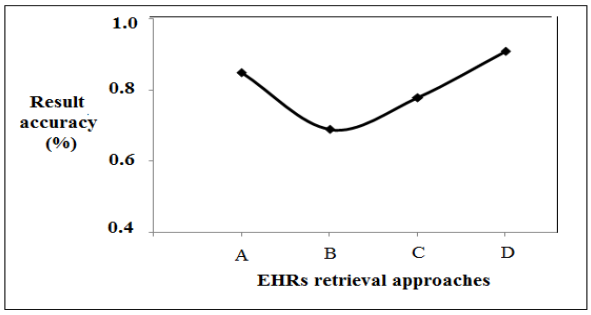


Fig. 3: Result Accuracy of EHRs retrieval approaches.

Efficiency measures observations

Efficiency measures namely execution time; memory cost and scalability are calculated by varying the number of EHRs for each transaction. Execution time is measured by varying the number the EHRs with hypertension dataset (D1) of 381 records, diabetes dataset (D2) of 403 records and meningitis dataset (D3) of 310 records. Dividing EHRs into equal partitions (Pi) namely,

P1 – <=100 EHRs
P2 – <=200 EHRs
P3 – <=300 EHRs
P4 – <=400 EHRs
P5 – >400 EHRs

IIOAB JOURNAL
ISSN: 0976-3104

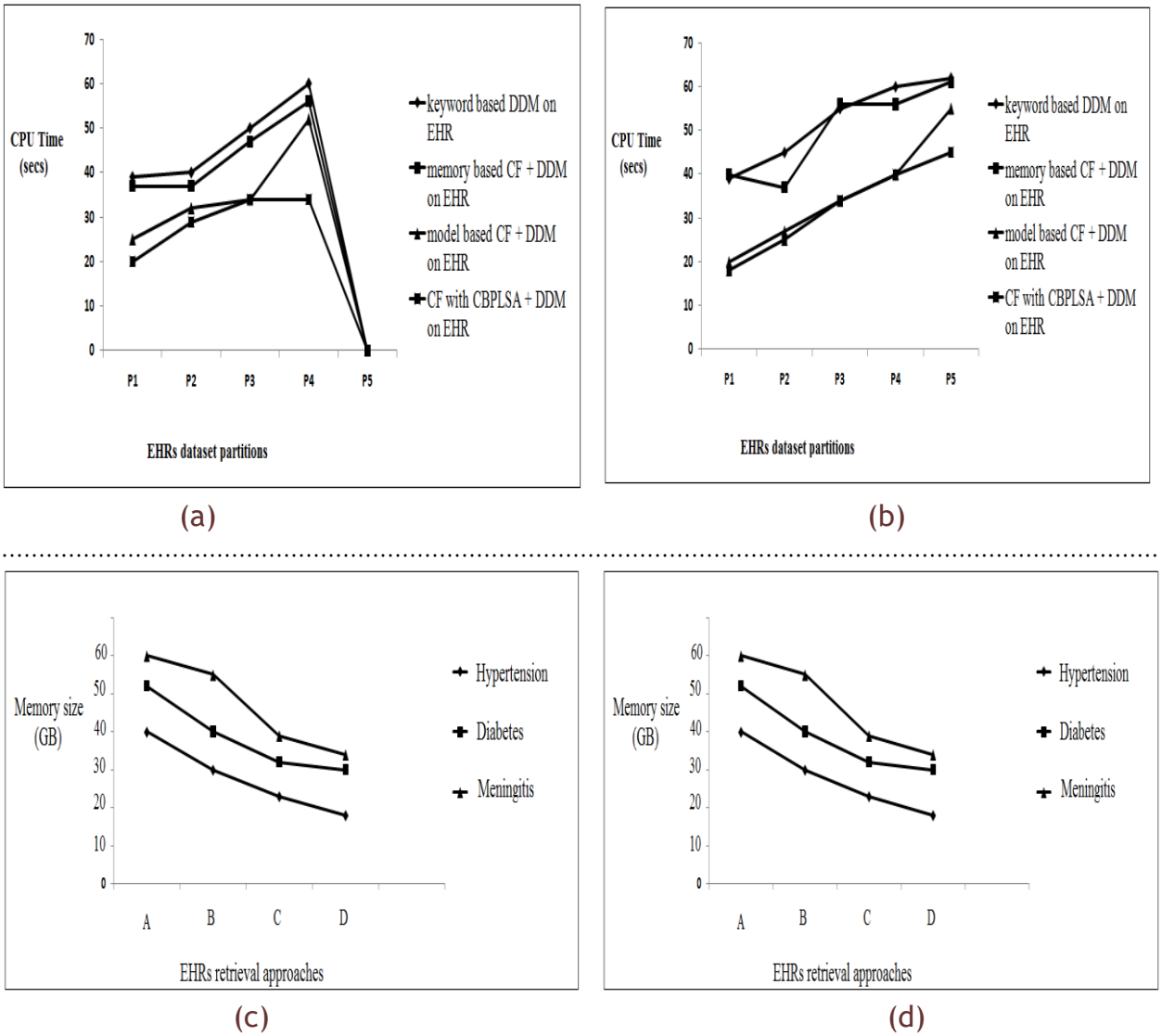


Fig. 4: Efficiency measures. (a) Dataset D1 vs CPU Time (b) Dataset D2 vs CPU Time (c) Dataset D3 vs CPU Time (d) EHRs retrieval approaches vs Memory size

From [Fig. 4(a), (b) and (c)] it is apparent that execution time for proposed CBPLSA algorithm of DDM on EHRs has minimum CPU runtime in the range of 20 to 30 secs for hypertension EHRs dataset (D1), 21 to 49 secs for diabetes EHRs dataset (D2) and 20 to 48 secs for meningitis EHRs dataset (D3).

From [Fig. 4(d)] it is obvious that memory cost for proposed CBPLSA algorithm of DDM on EHRs increases on increasing the number of EHRs datasets. However compared to other 3 state-of-arts EHRs retrieval approaches, proposed EHRs retrieval by memory-based CF and model-based CF (CBPLSA algorithm) of DDM on EHRs has minimum memory cost on increasing EHRs dataset.

Thus, the proposed architecture of CF using CBPLSA algorithm for DDM on EHRs shows an improved performance in-term of effectiveness and efficiency measures compared to other EHRs retrieval approaches. Further it shows that proposed CBPLSA algorithm minimizes computational complexity.

CONCLUSION

In this paper, architecture for EHR distributed mining with memory-based CF and model-based CF is proposed. Results shown proved that proposed architecture is intended for minimal computational complexity and memory overhead by maintaining clusters of old patients' EHRs. Also, performance evaluation of proposed CF using CBPLSA algorithm for DDM on EHRs shows improved precision, recall, F-measure and result accuracy values than individual memory-based CF on EHRs and model-based CF on EHRs. Further, focuses on finding appropriate CF framework for EHRs retrieval.

IIOAB JOURNAL
ISSN: 0976-3104

CONFLICT OF INTEREST
There is no conflict of interest.

ACKNOWLEDGEMENTS
None

FINANCIAL DISCLOSURE
None

REFERENCES

[1] Sawant V, Shah K. [2013] A review of Distributed Data Mining using agents. International Journal of Advanced Technology & Engineering Research (IJATER), 3(5):27-33.
[2] Cuzzocrea A. [2013] Models and algorithms for high-performance distributed data mining. Elsevier Journal of Parallel and Distributed computing, 73(93):281-283.
[3] Kargupta H, Kamath C, Chan P. [1999] Distributed and Parallel Data Mining: Emergence, Growth and Future Directions. Advances in Distributed Data Mining, (eds.), Hillol Kargupta and Philip Chan, AAAI Press, 407-416.
[4] Zaki MJ, Pan Y. [2002] Introduction: Recent Developments in Parallel and Distributed Data Mining. Springer Journal of Distributed and Parallel Databases, 11(2):123-127.
[5] Park BH, Kargupta H. [2002] Distributed Data Mining: Algorithms, Systems, and Applications. In. Data mining handbook. https://www.csee.umbc.edu/~hillol/PUBS/review.pdf
[6] Tsoumakas G, Vlahavas I. [2008] Distributed Data Mining, Encyclopedia of Data Warehousing and Mining. 2nd Edition John Wang (Ed.), Idea Group Reference, 709-715.
[7] Fu Y. [2001] Distributed Data Mining: An Overview. In: Newsletter of the IEEE Technical Committee on Distributed Processing,5–9.
[8] Khalilia M, Chakraborty S, Popescu M. [2011] Predicting disease risks from highly imbalanced data using random forest. BMC Med Inform Decis Mak. 11:51.
[9] Li Y, Bai C, Reddy CK. [2016] A distributed ensemble approach for mining health care data under privacy constraints. Inf Sci (Ny). 330:245-259..
[10] López-Nores M et al. [2012] Property-based collaborative filtering for health-aware recommender systems. 2011 IEEE International Conference on Consumer Electronics (ICCE). doi: 10.1109/ICCE.2011.5722619
[11] Davis DA et al, [2010] Time to CARE: a collaborative engine for practical disease prediction. Data Min Knowl Disc. 20:388–415.
[12] Zheng K, Mei O, Hanauer DA. [2011] Collaborative search in electronic health records. J Am Med Inform Assoc. 18(3):282-91.
[13] Koren Y, [2008] Tutorial on recent progress in collaborative filtering. In Proc. of the 2nd ACM Conference on Recommender Systems, 8-67.
[14] Su X, Khoshgoftaar TM. [2009] A Survey of Collaborative Filtering Techniques. Advances in Artificial Intelligence. doi:10.1155/2009/421425.
[15] Urmela S, Suresh Joseph K.[2015] An Effective Web Service Selection based on Hybrid Collaborative Filtering and QoS-Trust Evaluation. IJARCET, 3(3):69-78.

ARTICLE

RETHINKING CONSTRUCTION CORPORATE SOCIAL RESPONSIBILITY PRACTICES: CONSTRUCTION NEIGHBORHOOD

Keyvanfar, A.[1,2,3,4*], Khorami, M.[1], González Moya, Carmen[1], Cruz Cabrera, Marianela[1], Revelo Báez, C. Norma[1], Estupiñán, M.[1], Gamboa Pérez, F.[1], Pérez, C. E.[1], Alvansazyazdi, Mohammadfarid[5], Larco Benítez, Myrian A.[1], Martínez Serra, Juan Carlos[1], Avilés Díaz, Nelson Eduardo[1], Cadena Vallejo, R. E.[1], Paredes Lascano, I. S.[1], Granja Alencastro, Pablo R.[1], Torres Paucar, M. A.[1], Jácome Terán, William[1], Vivanco, Diana[1], García, Anggie[1], Shafaghat, A. [3,4]

[1]*Facultad de Arquitectura y Urbanismo, Universidad Tecnológica Equinoccial, Calle Rumipamba s/n y Bourgeois, Quito 170147, ECUADOR*

[2]*Jacobs School of Engineering, University of California, San Diego, 9500 Gilman Dr. La Jolla, CA 92093, UNITED STATES*

[3]*Faculty of Built Environment, Universiti Teknologi Malaysia, Skudai, 81310 MALAYSIA*

[4]*MIT-UTM MSCP Program, Institute Sultan Iskandar, Universiti Teknologi Malaysia, Skudai 81310, MALAYSIA*

[5]*Docente de la Facultad Ingeniería Ciencias Físicas y Matemática, Carrera Ingeniería Civil, Universidad Central del Ecuador, Quito 170129, ECUADOR*

ISSN 0976-3104

ABSTRACT

Background: *From a stakeholder management perspective, construction plays a big role in applying mitigation or adaptation solutions to eliminate the societal risk of development projects. Construction neighbourhood is a basic construction stakeholder with major and direct effect on work progress. Generally, corporate social responsibility (CSR) is the main business approach employed to regulate such a risk. It is while construction CSR is not properly considered the neighbourhood to address the risk.* ***Method:*** *Using the structure of CSR elements in the literature as a basis, this study critically reviews and argues CSR direction.* ***Findings:*** *The argument is outlined observed issues on CSR role, objective, stakeholder, and activity setting as well as comprehensive integration and CSR direction, and highlighted that CSR directions should be re-examined.* ***Significance:*** *This work intends to shed light on the field to redirect the efforts for better construction CSR practices.*

KEY WORDS
Construction Neighborhood, Socially Sustainable Construction Management, Construction Corporate Social Responsibility.

Received: 19 Mar 2018
Accepted: 29 Mar 2018
Published: 10 April 2018

***Corresponding Author**
Email:
alikeyvanfar@gmail.com
Tel.: +1-858-264-7085

INTRODUCTION

Construction is well known for its highly negative social impact. Neighborhood-related CSR activities have been practiced in construction since the Hammurabi times (nearly 3,700 years ago) [1] when a set of rules were established to protect affected people from urban development. Construction neighborhood is a community area and/or group directly influenced by construction activeness, including changes in the way of life, activities, asset values, jobs, friendships, and long- and short-term living plans. If such changes are mismanaged, the consequential negative effect may jeopardize a construction project (i.e., societal risk) (Jones, Comfort, and Hillier (2006) [2]). Construction fails to cover the risk properly. Overestimating social-carrying capacity always results in the irrational behavior of citizens, leading to a lose-lose situation. An example of such a situation is any case wherein resettlement (as a "firefighting strategy") is the only way forward, and construction must pay for it (Lin et al. (2017) [3]). Apart from its negative outcome, such as "social instability," the project loses profit margin.

CSR research aims to assist construction companies in considering advanced mitigation or adaptation strategies to optimize or minimize the negative social effect. Corporate governance, corporate citizenship, corporate accountability, and business ethics are other terms referring to the CSR concept (Duman, Giritli, and McDermott (2015) [4]). The UK seems more progressive in CSR implementation and research, although studies on this field are mainly from the USA, the UK, China, and Australia (Duman, Giritli, McDermott (2015) [4]). Although Duman, Giritli, and McDermott (2015) [4] defined CSR as difficult because it is a "multidimensional and nebulous concept," we consider it a simple, customizable, and profitable business strategy.

Studies report that CSR strategies and practices are growing in the construction industry [5] (Duman, Giritli, and McDermott (2015) [4]). However, construction neighborhood, as a basic construction stakeholder community, is not explored properly. The best study in this research area is [6], but it is incomprehensive, and the focus on the neighborhood is not as the main stakeholder. [Table 1] revisits the five most comprehensive review papers on CSR stakeholder, considering construction neighborhood as a stakeholder in addition to potential CSR concerns. Nearly in every report, construction neighborhood is not considered as a direct interest party.

IIOAB JOURNAL
ISSN: 0976-3104

Table 1: Reviews on construction neighborhood as a stakeholder

Stakeholders	Construction neighborhood as a stakeholder	Potential CSR concerns	Ref.
Employees, clients, future users, shareholders, suppliers, and particularly, local community and the public	Indirectly considered under public and local community	Emissions, effluents and waste, biodiversity, energy, water, indirect economic effects	[7]
Employees, customers, shareholders, creditors, suppliers and partners, environment and resource agencies, local communities, government, competitors, and NGOs.	Indirectly considered under project impact on the community	Protect the local environment, minimize safety hazards to the community, establish good communication channels with neighbors	[8]
Clients, Creditors (financial partners), employees, EPA, future users, governments, local community, NGOs, partners, shareholders, and suppliers/subcontractors	Indirectly considered under EPA and local community	EPA (water, land use, waste disposal, pollution emission, etc.), local community (project impact on the community, good communication channels with neighbors, etc.)	[6]
Governments (state, local, functional departments, etc.), society (media, community, NGOs, the public, etc.), and businesses (designers, supplier contractors, etc.)	Indirectly considered under business–society relations (stakeholder issues)	Harmonious relationships with local communities, employment from the local area (neighborhood)	[9]
Government, media, designers, and project legal personnel (contractor, supervisor, supplier, operator, public community, NGOs)	Indirectly considered under government, operator, media, and the public	Governments concerns (ensure transparent information disclosure), operators concerns (protect the local ecological environment), media concerns (express concern over the community and public requirements), the public concerns (maintain social stability, protect local community environment)	[3]

In the following sections, the study investigates CSR elements and lists a set of observed issues in the current CSR research and practices.

OBSERVED ISSUES

CSR has shown a great potential to mitigate social risks [18, 19]. Construction with an engineering nature [20-22] has been trying to localize CSR to its day-to-day activities. Several symptoms are associated with the poor adaptation and implementation of CSR in the construction industry. Duman, Giritli, and McDermott (2015) [4] reported that CSR is practiced more in developed countries than in developing ones and by large or international companies than SMEs (Lazarevic (2008)[10]). Also, CSR is only considered for mega projects (Duman, Giritli, and McDermott (2015) [4]).

From the extent of research, reports are merely observations, playing a passive role in the transformation of CSR to a problem-solving strategy. For example, existing research does not focus on proposing a CSR activity scientifically, put forward a system to engineer the implementation benefits, and then measure and validate the resulting stakeholder satisfaction/dissatisfaction.

The current study focuses on the observed issues associated with each CSR element. Five CSR elements are investigated in the construction literature [Table 2], including driving roles, objectives, stakeholders, activities, and integration plans. The arguments report issues on CSR role, objective, stakeholder, and activity setting as well as comprehensive integration.

Table 2: Elements of CSR in construction literature

Elements of CSR in construction	Description	Citation
Driving roles	Studies studied or briefly covered the driving role of CSR from the perspective of the construction industry	[4], [2], [11], [12], [13], [10], [14], [15]
Objectives	Studies attempted to cover CSR objectives from the perspective of the construction industry	[4], [15], [16], [17], [12], [13], [10], [9], [3]
Stakeholders	Studies attempted to identify or introduce CSR stakeholder in the construction industry	[4], [2], [11], [7], [12], [13], [14], [8], [6], [9], [3]
Activities	Studies attempted to identify or establish or briefly cover a set of CSR activities in the construction industry	[15], [4], [2], [11], [17], [7], [13], [14], [10], [8], [6], [9], [3]
Integration plans	Studies attempted to highlight the problem of CSR isolation, propose a solution or briefly cover CSR integration	[4], [2], [11], [9], [3]

IIOAB JOURNAL
ISSN: 0976-3104

CSR role-setting issues: CSR in construction has been in practice since the 1990s. CSR drivers for construction companies lean more toward social promotional strategy than a social problem-solving strategy. To illustrate, Barthorpe (2009)[14] indicated that being a "corporate citizen" add "credibility" and "enhance reputation." Evangelinos et al. (2016) [15] claimed to be responsive to non-financial aspects of performance. Lazarevic (2008)[10] indicated driving toward "local community protection and engagement." Based on his findings, the current practices focus more on engagement than protection to drive "effective and efficient building and constructing service as well effective management" of a business. Although Lazarevic (2008) [10]'s interpretation is better than the common one, the attention to problem-solving strategy remains missing.

Extant literature introduces CSR in construction as a strategy to increase the social caring capacity. Increasing social caring capacity is an attempt to induce people's tolerance to the negative impact they are facing for the sake of future benefits. Although no study exists to measure such success, the prevailing perception is unsustainable. Even if such idea is a forward-thinking method of role setting, the attempt remains a kind of "painkiller pills," that is, "killing the pain without trying to address the cause." In this situation, the construction industry is losing credit to develop an effective solution to rectify the problems while expanding CSR activities.

CSR objective-setting issues: Existing research fails to consider the functional analysis of CSR objectives as well as vision, mission, and target setting.

Schultz and Castello (2013) [16] classified characteristics of the instrumental, political-normative, and communication views on CSR. Lin et al. (2017) [12] defined three CSR perspectives, namely, profit, value, and relationship. Foreseeing the key role of functional analysis is possible for readers, but it is not considered as critical as it should be. Building social trust or branding by focusing on customers, NGOs or authorities is a common understanding of CSR practices. However, minimum attention is given to construction neighborhood, which should be basic. The minimum role of CSR that should be considered is the "must be there" social criterion (i.e., "basic operational" or "non-spoken" social criteria). Normally, perspective and paradigm in CSR focus on future value and profit, and this attention leads to forgetting the basics.

Moreover, construction companies perceive of a long-term return is frequently reported as a CSR implementation driver [4]. Defining CSR only as a long-term construction strategic management activity seems to be a common understanding. Thinking of CSR long-term return in terms of vision, mission, and target setting is another objective-setting problem. Construction is a short- medium-term profit-oriented industry. Thus, the orientation of CSR activities as a long-term strategy discourages considering the short-term duties. For example, CSR "…is a process and structure in which companies are directed and controlled to achieve long-term shareholders' goals concerning the interest of other stakeholders" [10]. As a result, construction neighborhood, which traditionally causes expense rather than creating profit, is forgotten. As such, long-term perception is only under the scope of international and mega project construction companies instead of SMEs for implementing CSR practices [4]. This limited scope is a side effect of long-term CSR objective setting.

CSR stakeholder-setting issues: CSR practices focus less on local community and neighborhoods. As evidence of this common practice on CSR, the phrase "outside stakeholders," also known as "non-shareholder stakeholders," is considered for construction neighborhood (Lin et al. (2017) [3]). Zhao et al. (2006) [6] investigated that CSR "…issues about clients, suppliers, partners, shareholders, and employees carry more weight than those concerning creditors, local community, environment, and competitors." Moreover, Jones, Comfort, and Hillier (2006) [2] stated that "some construction companies report their involvement in wider community initiatives and their charitable contributions to local communities."

Traditional thinking considers organized community group as construction stakeholder. In the old social power theory, "…most public people are not organizational but may be strongly influenced or guided by social organizations, such as the media, and NGOs" (Zeng et al. (2017) [9]). In the arena of social media, social network, and online communication, concurrent social power theory is different from the past. Lin et al. (2017) [12] indicated that this theory covers the ability of one stakeholder to influence the interest of another stakeholder (s). Only considering structured and organized or semi-organized community groups (with a direct benefit or formal legal binding and influence) is not promising. Construction neighborhood is even a non-organized community, but with the aid of communication technology, such community can be a bigger threat than an organized one.

To ensure the fulfillment of social demands, further research must identify the problems related to construction neighborhood. Development will then aid all parties involved in a construction project to construct a stable power construct in executing construction activities.

CSR activity-setting issues: Even the implementation of CSR activities serves as pertinent history in construction (Barthorpe (2009) [14]), the formal CSR implementation has been imported from social science discipline. That is why the scope of CSR solution is just limited to social science techniques with regard to adaptation and mitigation of construction issues. The earlier discussed social-carrying capacity role setting is one of the examples of such an adaptation of a social science method of thinking to

IIOAB JOURNAL
ISSN: 0976-3104

construction CSR. Available online literature presents a few pieces of evidence to this argument. Barthorpe (2009, Table 2) [14] reported a good set of examples, such as charity, social funds, and local recruitment, also known as stylish CSR activities. "Good communication with neighbors," use of "project newsletters," "providing up to three days paid leave per year to enable its employees to participate in approved local community projects," "work with disadvantaged members of communities," "race days, raft races, and marathon runs to raise money for local charities," "working with schools to promote children's life skills and it encourages employees to take part in charitable and community-based events" are a few of the successful examples in the literature (Jones, Comfort, and Hillier (2006) [2]). Zeng et al. (2017) [9] mentioned problems, such as "irrational behavior of the local community" and "unethical behavior local governments;" and possible solutions, such as community behavioral consultation to "think rationally" and "fair reporting by the media and proactive coordination by NGOs."

Although construction can be proud of all these ethical activities introduced by CSR to the industry, where are the business-oriented construction solutions? If construction neighborhood has been considered a key and basic stakeholder of CSR, then, we should have site planning CSR strategies, solutions, and activeness. The same applies to scheduling, resource management, and so on. As an expected result, Duman, Giritli, and McDermott (2015) [4] mentioned, "yet, many companies in this industry are having difficulties in integrating their social, ethical and environmental concerns into their operations and stakeholder interactions."

CSR comprehensive integration issues: Substantial efforts have been exerted on the professional integration of collaborative and concurrent engineering in construction. However, construction remains an isolated industry in strategic planning, and this status also affects CSR implementation. The relevant number of studies with the integration focus remains low, and only a few investigated the lack or the importance of CSR implementation (Duman, Giritli, McDermott (2015) [4]. In construction, we have several paralegal strategic efforts toward social impact management that may cover the social needs of construction neighborhood. However, the literature does not present adequate effort on the comprehensive integration of CSR to social impact assessment, sustainability, and socially sustainable assessment, construction risk assessment as well as parallel sustainable neighborhood assessment (SNA) discipline in urban planning and design research. Each discipline reports efforts toward sustainable social construction neighborhood under specific areas [Fig. 1]. Rethinking integration with the other disciplinary findings is a great improvement for these segmented research studies to boost the coverage and improve the scope of service. The ultimate consideration of the best CSR solutions is possible if it is based on the background of the effort of all parallel disciplines on the problems, causes, effects, and best practices. Such isolation partly explains why CSR remains not considered a problem-solving strategy with a direct financial return.

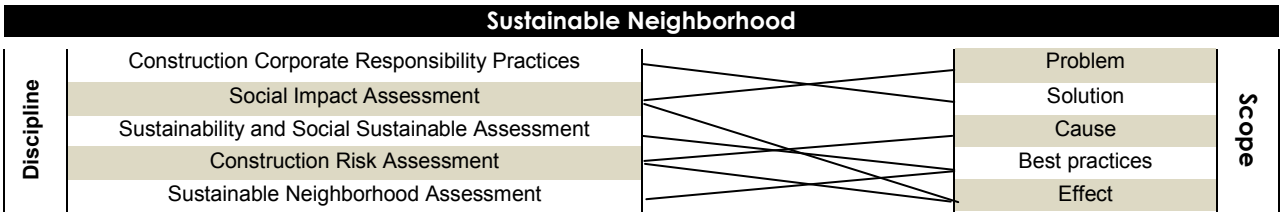


Fig. 1: Thematic coverage of construction neighborhood by other paralegal strategic efforts toward social impact management

In general, CSR has a great potential to boost academic–industry linkage, but this potential has been ignored. Integration with universities as an R&D stakeholder would help CSR, as well as universities, take a win-win step forward. In a way, CSR is the only communication mode of construction companies. CSR is a concept that the industry intends to adopt, so boosting this practice by reporting problems and possibly applying solutions is crucial. CSR reports may provide researchers the means to investigate demand-driven industrial problems. In addition, by providing CSR research funds to solve company or industry problems as well as promoting R&D, is another possible way. From the R&D side, financial contributions and the recognition of R&D need are both extremely helpful to finally put into practice strong R&D habit, culture, and experiences in the construction industry.

CONCLUSION

This effort has been made to redirect the attention of researchers on the future development of CSR literature to construction neighborhood. CSR elements have been investigated through literature review. Relevant construction neighborhood CSR implementation issues have been discussed under each element. These issues are on CSR role, objective, stakeholder, and activity setting as well as comprehensive integration. This study has been reported with the aim to inspire the rethinking of construction CSR practices. One of the most significant solutions proposed is the integration of CSR with the R&D sector. Possible information that can contain in the CSR report, as well as possible CSR financial contribution of construction companies, can boost R&D ecosystem, to help solve the industry problem,

ISSN: 0976-3104

guarantee a financial return to expanding companies. This report has consistently inspired readers to consider CSR as a business solution strategy with possible short-term return rather than merely for branding and seeking a long-term return.

CONFLICT OF INTEREST

The authors do not have any conflict of interest.

ACKNOWLEDGMENTS

We would like to thank the anonymous referees for their helpful suggestions that substantially improved this article. This report presents the research conducted with support from the UTE as a part of an effort (over 3,000 man/hours of research to collect and process approximately 35,000 data) in the Ecuador construction industry. Our mission is to provide demand-driven results on sustainable solutions to the current construction problems in the construction industry. We trust the current collaborative research output results in a win-win scenario for the future sustainable construction industry. To this end, we are honored to receive the humble and dedicated support of Dr. Ricardo Hidalgo Ottolenghi, the director of UTE, as well as Dr. Madam Myriam Larco, dean of the UTE Faculty of Architecture and Urbanism, whose forward-thinking substantially contributed to the success of this program.

FINANCIAL DISCLOSURE

The research was supported by Universidad Tecnológica Equinoccial (UTE), Ecuador

REFERENCES

[1] Asongu JJ. [2007] Strategic corporate social responsibility in practice. Greenview Publishing Co.

[2] Jones P, Comfort D, Hillier D. [2006] Corporate social responsibility and the UK construction industry. Journal of Corporate Real Estate, 8(3): 134-150.

[3] Lin H, Zen, S, Ma H, Zeng R, Tam VW. [2017] An indicator system for evaluating megaproject social responsibility. International Journal of Project Management.

[4] Ulutaş Duman D, Giritli H, McDermott P. [2016] Corporate social responsibility in construction industry: A comparative study between UK and Turkey. Built Environment Project and Asset Management, 6(2):218-231.

[5] Petrovic-Lazarevic S. [2008] The development of corporate social responsibility in the Australian construction industry. Construction Management and Economics, 26(2): 93-101.

[6] Zhao ZY, Zhao XJ, Zuo J, Zillante G. [2016] Corporate social responsibility for construction contractors: a China study. Journal of Engineering, Design and Technology, 14(3):614-640.

[7] Weisheng Lu, Meng Ye, Roger Flanagan, Kunhui Ye. [2015] Corporate social responsibility disclosures in international construction business: trends and prospects. Journal of Construction Engineering and Management 142(1): 04015053.

[8] Zhao, Zhen-Yu, Xiao-Jing Zhao, Kathryn Davidson, Jian Zuo. [2012] A corporate social responsibility indicator system for construction enterprises. Journal of cleaner production 29: 277-289.

[9] Hanyang Ma, Saixing Zeng, Han Lin, Hongquan Chen, Jonathan J Shi. [2017] The societal governance of megaproject social responsibility. International Journal of Project Management 35(7): 1365-1377.

[10] Sonja Petrovic-Lazarevic. [2008] The development of corporate social responsibility in the Australian construction industry, Construction Management and Economics, 26(2): 93-101.

[11] Martin L, Lim BTH. [2017] Linking corporate social responsibility and organizational performance in the construction industry. Construction management and economics 35(3): 90-105.

[12] Lin X, Ho CM, Shen GQ. [2017] Research on corporate social responsibility in the construction context: a critical review and future directions. International Journal of Construction Management, 1-11.

[13] Loosemore M, Lim BTH. [2018] Mapping corporate social responsibility strategies in the construction and engineering industry. Construction Management and Economics 36(2): 67-82.

[14] Barthorpe S. [2010] Implementing corporate social responsibility in the UK construction industry, Property Management, 28 (Iss 1): 4 – 17

[15] Konstantinos E et al. [2016] Exploring the status of corporate social responsibility disclosure in the UK building and construction industry. International Journal of Global Environmental Issues 15(4): 377-399.

[16] Schultz F, Castelló I, Morsing M. [2013] The construction of corporate social responsibility in network societies: A communication view. Journal of business ethics, 115(4): 681-692.

[17] Jiang W, Wong JKW. [2016] Key activity areas of corporate social responsibility (CSR) in the construction industry: a study of China. Journal of cleaner production 113: 850-860.

[18] Banerjee SB. [2008] Corporate social responsibility: The good, the bad and the ugly. Critical sociology, 34(1): 51-79.

[19] Windsor D. [2001] The future of corporate social responsibility. The international journal of organizational analysis, 9(3): 225-256.

[20] Mohammadhassani M et al. [2013] Identification of a suitable ANN architecture in predicting strain in tie section of concrete deep beams. Struct. Eng. Mech, 46(6): 853-868.

[21] Safa M et al. [2016]. Potential of adaptive neuro fuzzy inference system for evaluating the factors affecting steel-concrete composite beam's shear strength. Steel and Composite Structures, 21(3): 679-688.

[22] Kafi MA et al. [2014]. Evaluating the seismic performance of off-centre bracing system with circular element in optimum place. International Journal of Steel Structures, 14(2):293-304.

ARTICLE

DEVELOPMENT OF LOOP DESTRUCTIVE NETWORK ANALYSIS

Keyvanfar, A.[1,2,3,4*], Khorami, M.[1], Avilés Díaz, Nelson Eduardo[1], Cruz Cabrera, Marianela[1], González Moya, Carmen[1], Martínez Serra, Juan Carlos[1], Gamboa Pérez, F[1], Revelo Báez, C. Norma[1], Estupiñán, M.[1], Cadena Vallejo, R. E.[1], Granja Alencastro, Pablo R.[1], Alvansazyazdi, Mohammadfarid[5], Paredes Lascano, I. S.[1], Torres Paucar, M. A.[1], Jácome Terán, William[1], Larco Benítez, Myrian A.[1] ,García, Anggie[1], Vivanco, Diana[1], A. Shafaghat[3,4]

[1]*Facultad de Arquitectura y Urbanismo, Universidad Tecnológica Equinoccial, Calle Rumipamba s/n y Bourgeois, Quito 170147, ECUADOR*

[2]*Jacobs School of Engineering, University of California, San Diego, 9500 Gilman Dr. La Jolla, CA 92093, UNITED STATES*

[3]*Faculty of Built Environment, Universiti Teknologi Malaysia, Skudai 81310, MALAYSIA*

[4]*MIT-UTM MSCP Program, Institute Sultan Iskandar, Universiti Teknologi Malaysia, Skudai 81310, MALAYSIA*

[5]*Docente de la Facultad Ingeniería Ciencias Físicas y Matemática, Carrera Ingeniería Civil, Universidad Central del Ecuador, 170129, Quito, ECUADOR*

ISSN 0976-3104

ABSTRACT

***Background**: Research on risk network analysis has not developed a simple method that is based on the destructive network function of nodes. **Objective**: Hence, we propose loop destructive network analysis (LDNA), a method of identifying the most influential relation and node to break (destruct) a major number of loops. For this purpose, general connectivity is reduced with minimal effort. **Case study**: LDNA is applied in the connective network of construction corporate social responsibility (CSR) implementation risks. The process and result are showcased and discussed to deliver a clear understanding on LDNA. **Significance**: This method will aid the success of resource planning in systematic risk reduction.*

KEY WORDS
Risk analysis
Dynamic modelling
Network of risks

INTRODUCTION: MODEL DEVELOPMENT

Risk analysis [1–7] and dynamic modeling [8–11] have been studied to show construction stakeholders the best identification and analysis approaches for risk networks. These networks contain negative loops, and possible issues resulted from negative loops in a continuous and stronger risk effect. Studies have focused on identifying and rectifying risk of nodes and reducing the effect of one node of risk to another. However, research on risk network analysis has not developed a simple method that is based on the destructive network function of nodes and relationships in loops of negative effects. In general, two perspectives that are based and depend on the nature of phenomena are used to destroy major loops in networks, namely, node- and relation-based loop destruction. A simplified model proposal framework for both perspectives is described below.

Step 1. Identifying loops: The loops are mapped and are classified on the basis of the number of nodes involved. For example, we should seek "n−1" type loops in the network of "n" node (e.g., 2-node type, 3-node type, ..."n"-node type of loops).

Step 2. Identifying the most loop-destructive-single node(s) and/or relationship(s): The destructive factor (DF) of each node and/or relation considered dividing the number of loops that the node involved in the total number of loops.

Step 3. Identifying the best sequential partner(s) nodes and/or relations most loop-destructive-single node(s): The highest DF that combines the minimum number of nodes and /or relations that ensures the maximum number of alternative solutions is reached. The result is introducing optimum node and/or relation combinations to destroy all the loops in the networks and to reach a full DF of 1.

CASE STUDY

[Fig. 1] shows an observed conceptual network of issues in implementing construction corporate social responsibility (CSR). The network is adopted from Keyvanfar et al. (2018) [12]. Every node in [Fig.1] influences and is influenced by another set of nodes. The network possesses several negative infinitives loops. We must eliminate effective nodes in the network to optimize the reduction of the negative effect of the loops. The target problem is eliminating the minimum number of nodes to destroy the loops only. The proposed step-by-step node-based LDNA is discussed below.

Received: 28 March 2018
Accepted: 5 April 2018
Published: 15 April 2018

***Corresponding Author**
Email:
alikeyvanfar@gmail.com
Tel.: +1-858-264-7085

IIOAB JOURNAL
ISSN: 0976-3104

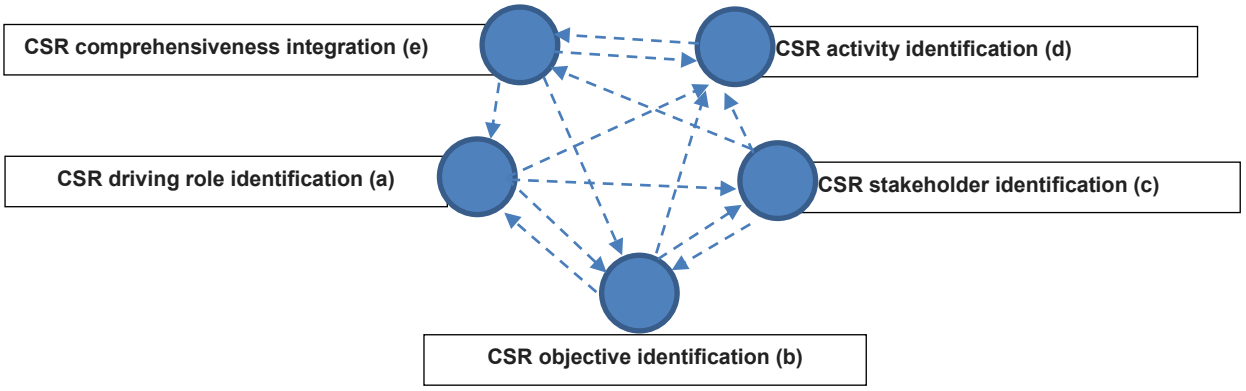


Fig. 1: Connective network of CSR implementation risks

Step 1: The loops are identified. Two-, three-, four-, and five-node loops are present. In [Fig.2], we classify the loops to introduce the investigated loop in the studied network. Fifteen loops are investigated, of which three, four, five, and three loops are two-, three-, four-, and five-node types, respectively.

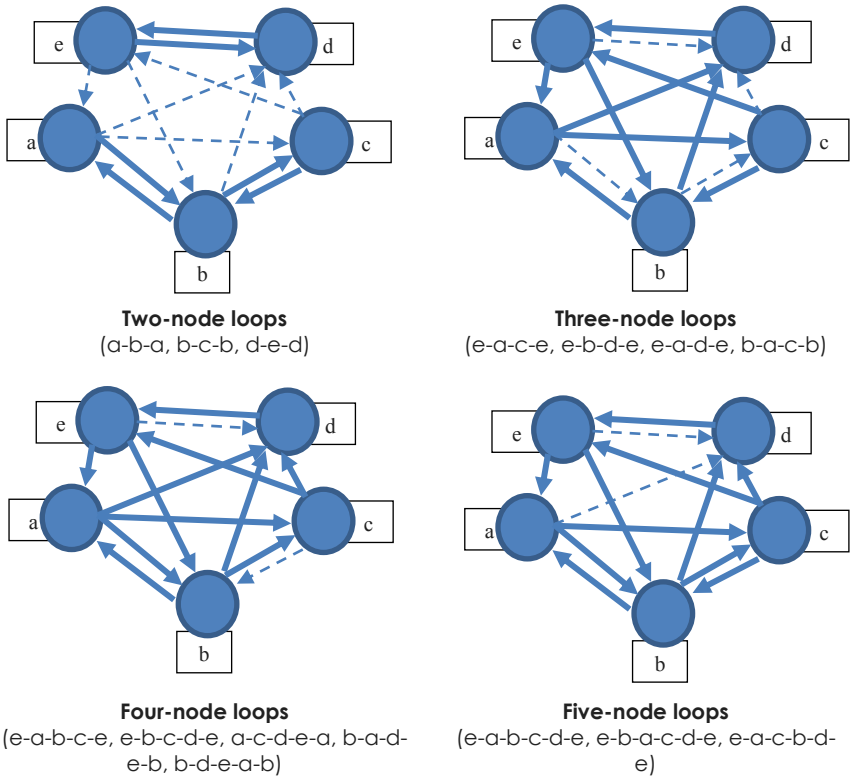


Fig. 2: Loop mapping of the connective network of CSR implementation risks

Step 2: The most involved nodes in the different loops are investigated. The DF for each node is calculated [Table 1]. The DF of each node considers dividing the number of loops that the node involved in the total number of loops. The DFs of nodes (e) and (c) of 0.8 and 0.6 are the highest and lowest, respectively.

Table 1: DF calculation based on investigating most repeated node in network loops of CSR issues

	Node	Frequency	DF
a	CSR driving role identification issues	11	0.733
b	CSR objective identification issues	11	0.733
c	CSR stakeholder identification issues	9	0.6
d	CSR activity identification issues	10	0.667
e	CSR comprehensiveness integration issues	12	0.8

IIOAB JOURNAL
ISSN: 0976-3104

Step 3: The best partners of the node (e) are investigated.
First, the DFs of every combination of the node (e) with other nodes are calculated [Table 2]. Alternative nodes (a), (b), (c), and (d) are available for possible partnership with (e) (For example, the partnership of (e) and (a) is accumulation of DF of (e) and (a) without double-counting the shared loops in which (e) and (a) both exist.). Within these alternatives, combinations of (e) with (a), (b), or (c) produce the maximum DF. (d) cannot be an effective partner because the DF of its combination with (e) is low and is even the same as that of (e) alone. Given that the DF of two-node combinations is not satisfactory (i.e., 1), we identify possible three-node combinations.

The DF of every combination of other nodes with (e-a) is listed in [Table 3]. The DF of the combination of (e-a) with (b) or (c) results in a satisfactory level of 1. The same result is obtained for the combination of (e-c), and (a) or (b) and for that of (e-b) and (a), or (c), or (d). The combination of (e-b) presents three alternatives to find the best third partner, whose partnership is preferred for destructive efforts to those of other alternatives introduced in step 2.

This will guide us to introduce the combination of node (e), and (b), and (a), or (c), or (d) as the result of the current LDNA case study, where (e) is the most loop-destructive node.

Table 2: DF calculation to establish the best two-node partnerships with most loop-destructive node

Node combination	Frequency	DF	Node combination	Frequency	DF
a and b	14	0.933	b and d	14	0.933
a and c	13	0.867	b and e	14	0.933
a and d	14	0.933	c and d	14	0.933
a and e	14	0.933	c and e	14	0.933
b and c	13	0.867	d and e	12	0.8

Table 3: DF calculation to establish best three-node partnership with most loop-destructive node

Node combination	Frequency	DF	Node combination	Frequency	DF
a and b and c	14	0.933	a and d and e	14	0.933
a and b and d	15	1	b and c and d	15	1
a and b and e	15	1	b and c and e	15	1
a and c and d	15	1	b and d and e	15	1
a and c and e	15	1	c and d and e	14	0.933

Addressing comprehensive CSR integration risks [node (e)] will help address network of risks and is the most efficient target to consider. Such findings are important because researchers can reconsider the future direction of this body of knowledge and effectively eliminate the problems of CSR implementation.

SIGNIFICANCE

LDNA is partially showcased in this case study to introduce its steps and logic of approach to readers. The proposed method is only for small human-based decision-making group discussions. The method can significantly support relevant decision-making in reducing the negative loops of a network of risk. Although we use construction risk as the impetus for this research, the method is appropriate for general applications, and it will be introduced in upcoming manuscripts.

CONFLICT OF INTEREST

The authors do not have any conflict of interest.

ACKNOWLEDGMENTS

We would like to thank the anonymous referees for their helpful suggestions, which substantially improved this article. This research was conducted with support from the Universidad Tecnológica Equinoccial (UTE) (over 3,000 labor-hours of research for collecting and processing approximately 35,000 data points) as part of an initiative in the Ecuador construction industry. Our mission is to provide demand-driven results on sustainable solutions to current problems in the construction industry. We trust that the current collaborative research results in a mutually beneficial scenario for the future sustainable construction industry. Thus, we are honored to receive the valuable and dedicated support of Dr. Ricardo Hidalgo Ottolenghi, director of UTE, and Dr. Madam Myriam Larco, dean of the UTE Faculty of Architecture and Urbanism, whose foresight substantially contributed to the success of this program.

FINANCIAL DISCLOSURE

The research was supported by the UTE (Ecuador)

REFERENCES

[1] Kangari R, Riggs LS. [1989] Construction risk assessment by linguistics. IEEE transactions on engineering management, 36(2): 126-131.

[2] Tah JHM, Carr V. [2000] A proposal for construction project risk assessment using fuzzy logic. Construction Management & Economics, 18(4):491-500.

[3] Carr V, Tah JHM. [2001] A fuzzy approach to construction project risk assessment and analysis: construction project risk management system. Advances in engineering software, 32(10-11):847-857.
[4] Zeng J, An M, Smith NJ. [2007] Application of a fuzzy based decision making methodology to construction project risk assessment. International journal of project management, 25(6): 589-600.
[5] Zavadskas EK, Turskis Z, Tamošaitiene J. [2010] Risk assessment of construction projects. Journal of civil engineering and management, 16(1):33-46.
[6] Kokangül A, Polat U, Dağsuyu C. [2017] A new approximation for risk assessment using the AHP and Fine Kinney methodologies. Safety science, 91:24-32.
[7] Ning X, Qi J, Wu C. [2018] A quantitative safety risk assessment model for construction site layout planning. Safety science, 104:246-259.
[8] Nasirzadeh F, Nojedehi P. [2013] Dynamic modeling of labor productivity in construction projects. International journal of project management, 31(6): 903-911.
[9] Javed AA, Zhan W, Pan W. [2018] A System Dynamics Framework of Drivers and Constraints to Enhancing Productivity of the Hong Kong Construction Industry. In Proceedings of the 21st International Symposium on Advancement of Construction Management and Real Estate (pp. 117-127). Springer, Singapore.
[10] Tesfaye E, Berhan E, Kitaw D. [2018] A Comprehensive Literature Review on Construction Project Risk Analysis. Global Business Expansion: Concepts, Methodologies, Tools, and Applications: Concepts, Methodologies, Tools, and Applications, 289.
[11] Chen X, Lu WS. [2018] Scenarios for Applying Big Data in Boosting Construction: A Review. In Proceedings of the 21st International Symposium on Advancement of Construction Management and Real Estate (pp. 1299-1306). Springer, Singapore.
[12] Keyvanfar A, Khorami M, Gonzalez Moya, et al. [2018] Rethinking Construction Corporate Social Responsibility Practices: Construction Neighborhood, IIOABJ, 9(3): 9-13.

ARTICLE

SEASONAL VARIATION DETERMINATION OF PHYTOPLANKTON DENSITY IN AQUATIC SYSTEM

Waseem A. Gad*

Central Lab., Kafr-El-Sheikh Water and Wastewater Company, Kafr-El-Sheikh, EGYPT

ISSN 0976-3104

ABSTRACT

Algae might be one of the biggest causes of fresh water problems in Kafr El-Sheikh the governorate, Egypt; where phytoplankton density increases in certain periods of the year for many reasons and negatively affects the quality of fresh water used as resources by drinking water purification plants in terms of changing color, taste and smell. This study aimed to monitor the seasonal variation of fresh water phytoplankton density in front of intakes of three water purification plants located on Nile River in terms of counting and classification of algae as well as treated water phytoplankton density for determination of raw water situation and evaluation of water purification plants efficiency to remove algae. Results of this study showed that phytoplankton density increased in spring and summer seasons more than autumn and winter, while efficiency of algal removal exceeded 90 % for the three tested plants; finally this study provided a classification and an identification of certain types of dominant species of algae found in water

KEY WORDS
Pollution, Diatoms, Green Algae, Blue Green Algae, Efficiency Evaluation

Received: 08 November 2017
Accepted: 12 December 2017
Published: 17 April 2018

***Corresponding Author**
Email:
withyou_2010@yahoo.com
Tel.: +201015119474

INTRODUCTION

There is no life without water as water is one of the most essential necessities of life on earth. All countries of the world have turned to the importance of water quality because of its great impact on human health. Production of safe potable water is of great importance in precaution of incidence of many water-born diseases. Phytoplankton in particular species belonging to cyanobacteria (blue-green algae) are very significant because of their impact on the quality of water Because of the secretion of some substances that change the color, taste and smell of water as well as the secretion of some toxins that affect human health and cause diseases, they also have an impact on many environmental factors related to the quality of water as well as they may negatively affect the process of purification itself as algae species belonging to diatoms might cause filters clog because of their silicon frustules. [1].

Aquatic systems all over the world have been affected by the excessive discharge of many types of pollutants, causing phytoplankton blooms and affecting the environmental characteristics of those systems in many ways [2, 3, and 4]. And with the civilization leap in all countries of the world; a specific need have been originated to analyze and follow up the situation of these water resources to determine the adequacy of water purification processes and changes that may arise in certain periods as a result of pollution and monitor these changes for the sake of prediction of crises before their occurrence and work to find solutions to avoid disasters that may occur and affect human health and safety. this perception prompted researchers around the world to carry out many studies and statistical analyses to know about the reasons for increasing of algal density in raw water at certain periods and link between this increase and other various environmental factors to determine the reasons of this increase and follow its impact on the status of water and the process of purification in order to find appropriate solutions [5, 6 and 7].

Water resources pollution is considered one of the most critical problems affecting Nile River in Egypt. Pollution of Nile River has grown in the last few years because of increases in population; several new irrigated agriculture projects which have been increased with the expansion of agricultural plantation causing a huge increase of the quantity of agricultural drainage water discharging into Nile water and other activities along the Nile. As a result the dilution capacity of the Nile River system will diminish at the same time that the growth in industrial capacity is likely to increase the volume of pollutants discharged to the Nile. Other sources of pollution including sewage discharge regarding the increase of population [8].

The phenomenon of algae bloom on the River Nile and its tributaries became common in the last six years. Increased algae bloom negatively affects purification plants which have not been constructed to deal with high density of algae. Increased density of phytoplankton can cause serious problems in purification plant basins considering the accumulation of algae on their walls, frequently clogging of filters, affecting the purification process and increase the dose of coagulant and chlorination which cause financial increases. Consecutive Clogging of filters results in more back-wash water to be used, consuming more power and reducing water supply level to clients. Phytoplankton density increase also affects potable water quality through increasing alkalinity, pH, slightly and turbidity. As well as increasing clients complain from bad smell and odor starting from grassy, spicy, musty, and fishy to septic water [9].

MATERIALS AND METHODS

Sampling Site Description: Water samples were collected seasonally for one year (2016) from Intakes of three water purification plants located on Rosetta branch of Nile River. Water treatment plants are: 1) Mahalet Abo-Ali drinking water purification plant. 2) Fowah drinking water purification plant. 3) Mettobus drinking water purification plant. [Table 1] and [Fig.1] showed the three sites of sampling. Samples were

IIOAB JOURNAL
ISSN: 0976-3104

taken manually 30 cm under the surface of the water and transferred to Laboratory within an ice-box according to standard methods of water and wastewater [10].

Counting and Identification of algal count: A volume of 10 ml of each sample of raw and treated water was centrifuged for 10 min at 4000 rpm by Centrifuge, the supernatant was discarded, and then the pellet was dissolved in 1 ml of saline water, which was transferred to Sedgwick–rafter counting chamber that was covered by a slide cover. Algae were counted using inverted microscope fitted with colour video camera attached to a TV monitor. The count was carried out using the TV monitor; Diatoms, green algae, blue green algae and total algal count were recorded for each sample as showed by [10].

Table 1: Representing the sites of sampling and their coordinates

Site of Sampling	Latitude	Longitude
Mahalet Abo Ali Purification plant	31° 6'7.92"N	30°41'53.57"E
Fowah Purification plant	31°12'16.24"N	30°34'12.69"E
Mettobus Purification plant	31°16'56.68"N	30°31'28.43"E

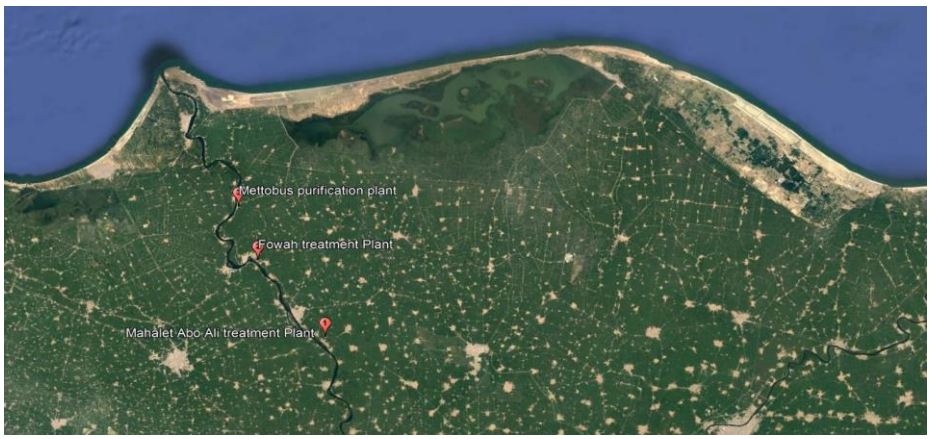


Fig. 1: Representing the three sampling sites and their locations on Rosetta Branch of Nile River.

RESULTS

This study monitored seasonal variations of algae densities in raw and treated water collected from three water purification plants in Kafr El-Sheikh Governorate which their intakes are located on Rosetta Branch of Nile River. Samples were investigated for the presence and count of Diatoms, Green Algae, Blue Green Algae and Total Algal Count around four seasons of one year (2016). This study also tried to evaluate the performance of algal removal of the tested purification plants.

Results of this study indicated that Total Algal count, Diatoms, Green Algae and Blue Green Algae increased in summer and autumn rather than winter and spring seasons with regard to Mahlet Abo Ali water purification plant. The highest Total Algal count (10860 org. /ml.) was recorded in summer 2016. The lowest Total Algal Count (7200 org. /ml.) was recorded in spring 2016 with regard to raw water samples. With reference to treated water samples; results revealed great removal ratio along the four seasons, The lowest value (650 org. /ml.) was recorded in winter 2016 while the highest value was recorded (840 org. /ml.) in summer 2016. [Table 2]and [Fig. 2] represented counts of Diatoms, Green and Blue Green algae including the Total Algal Count of raw and treated water samples of Mahlet Abo Ali water purification plant. Concerning efficiency evaluation of Algae removal of Mahlet Abo Ali water purification plant; results showed similar pattern over the four seasons. Removal rate was recorded 90.4 %, 90.2 %, 92.2 % and 91.8 % in winter, spring, summer and autumn 2016 respectively. The highest removal rate was recorded in summer 2016.

Concerning Fowah water purification plant; Results indicated that Total Algal count, Diatoms and Green Algae densities increased in summer and autumn rather than winter and spring seasons with regard to raw water samples. Except for Blue Green Algae count; results showed that their count increased in summer season followed by winter season. The highest Total Algal count (11450 org. /ml.) was recorded in summer 2016.The lowest Total Algal Count (7850 org. /ml.) was recorded in winter 2016 with regard to raw water samples. With reference to treated water samples; results revealed satisfactory removal ratio all over four seasons, The lowest value (580 org. /ml.) was recorded in spring 2016 while the highest value was recorded (800 org. /ml.) in summer 2016.[Table 2] and [Fig. 3] represented counts of Diatoms, Green and Blue Green algae including the Total Algal Count of raw and treated water samples of Fowah water purification plant. Concerning efficiency evaluation of Algae removal of Fowah water purification plant; results showed similar removal ratio all over the four seasons. Removal rate was recorded 91.8 %, 92.7 %, 93.0 % and 92.8 % in winter, spring, summer and autumn 2016 respectively. The highest removal rate was recorded in summer 2016.

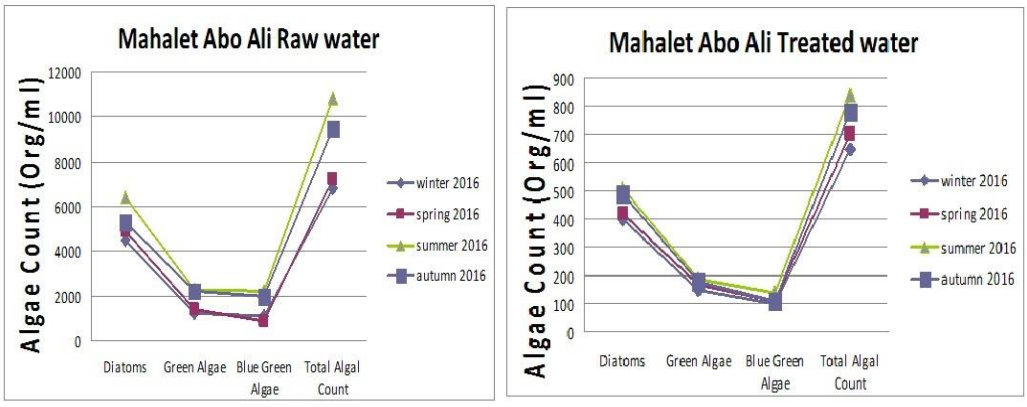
IIOAB JOURNAL
ISSN: 0976-3104

As for Mettobus water purification plant; Results showed that Total Algal count, Diatoms and Green Algae densities increased in summer season rather than any other season concerning raw water samples. The highest value of Blue Green Algae count was recorded in spring season. The highest Total Algal count (11100 org. /ml.) was recorded in summer 2016, while the lowest Total Algal Count (7000 org. /ml.) was recorded in winter 2016 with regard to raw water samples. Concerning treated water samples; results revealed acceptable removal rate around the year, The lowest value (450 org. /ml.) was recorded in winter 2016 while the highest value was recorded (740 org. /ml.) in summer 2016. [Table 2]and [Fig.4] represented counts of Diatoms, Green and Blue Green algae including the Total Algal Count of raw and treated water samples of Mettobus water purification plant. Efficiency evaluation of Algae removal of Mettobus water purification plant showed adequate removal rate around the year which was recorded 93.6 %, 93.9 %, 93.3 % and 93.7 % in winter, spring, summer and autumn 2016 respectively. The highest removal rate was recorded in spring 2016 in this case.

Efficiency evaluation of the three water purification plants under investigation showed satisfactory and accepted capabilities of algae removal over the course of four seasons of 2016 representing the period of study. [Fig.5] indicated the efficiency evaluation of the three plants around the year based on plants capability of algal removal throughout the four seasons of the period of study.

Table 2: representing seasonal variations of Total Algal Count, Diatoms, Green Algae and Blue Green Algae Counts recorded for raw and treated water samples of Mahalet Abo Ali, Fowah and Mettobus water purification plants

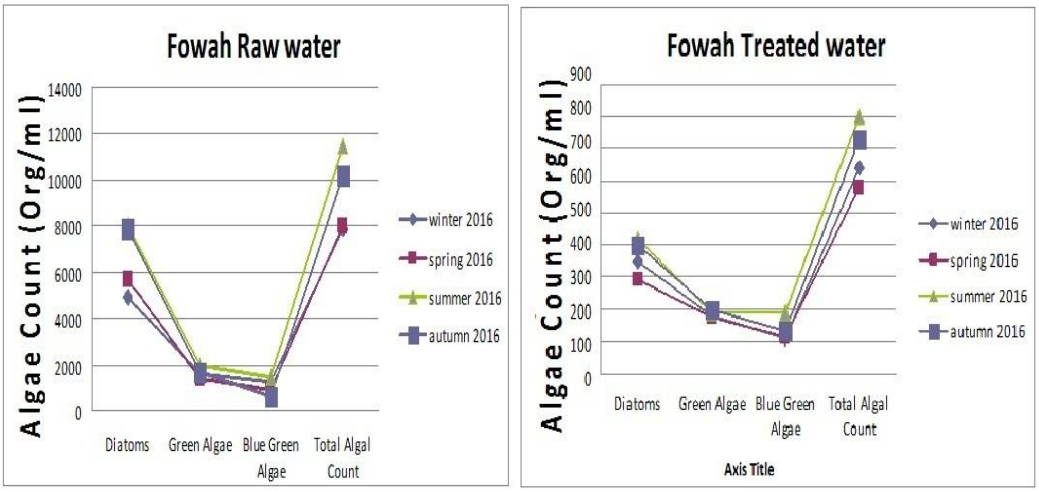
		Raw water				Treated water			
		winter 2016	spring 2016	summer 2016	autumn 2016	winter 2016	spring 2016	summer 2016	autumn 2016
Mahalet Abo Ali water purification plant	Diatoms	4500	4900	6400	5300	400	420	510	490
	Green Algae	1200	1400	2250	2200	150	170	190	180
	Blue Green Algae	1100	900	2210	2000	100	110	140	110
	Total Algal Count	6800	7200	10860	9500	650	700	840	780
Fewwah water purification plant	Diatoms	4900	5700	8000	7850	350	290	420	400
	Green Algae	1650	1400	2000	1700	180	175	190	200
	Blue Green Algae	1300	900	1450	650	110	115	190	130
	Total Algal Count	7850	8000	11450	10200	640	580	800	730
Mettobus water purification plant	Diatoms	4200	4850	7500	6350	280	300	480	430
	Green Algae	1650	1910	2450	1730	120	180	130	140
	Blue Green Algae	1150	1740	1150	1720	50	40	130	50
	Total Algal Count	7000	8500	11100	9800	450	520	740	620



A. B.

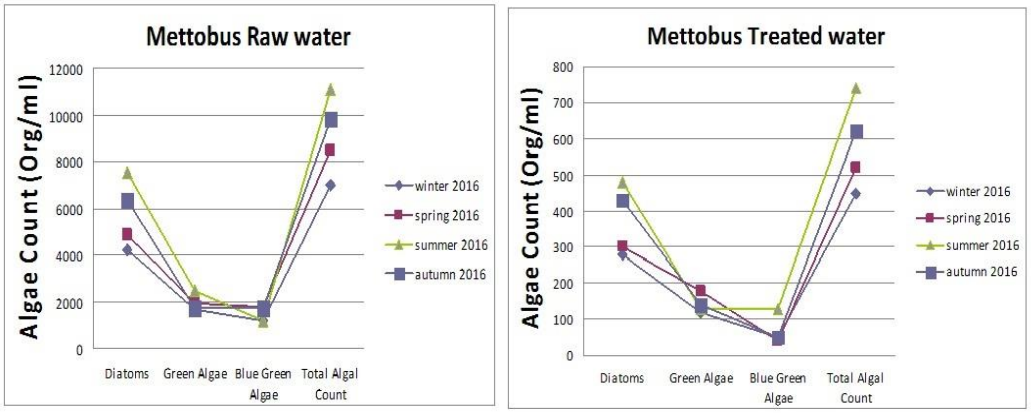
Fig.2: Representing Seasonal variation of phytoplankton density in raw and treated water regarding Mahalet Abo Ali water purification plant; where, A. representing raw water and B. representing treated water.

IIOAB JOURNAL
ISSN: 0976-3104



C. D.

Fig.3: Representing Seasonal variation of phytoplankton density in raw and treated water regarding Fowah water purification plant; where, C. representing raw water and D. representing treated water.



E. F.

Fig. 4: Representing Seasonal variation of phytoplankton density in raw and treated water regarding Mettobus water purification plant; where, E. representing raw water and F. representing treated water.

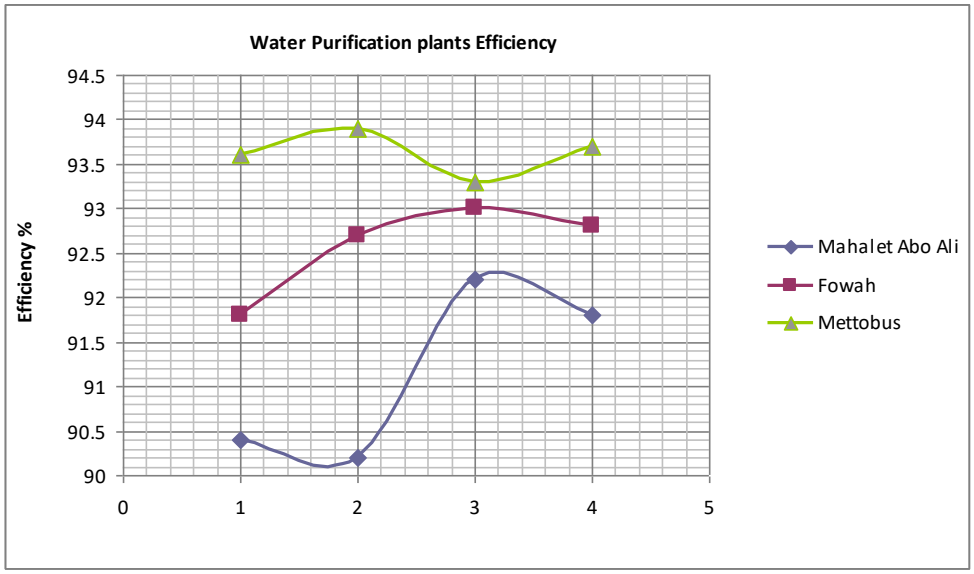


Fig.5: Representing Efficiency evaluation of the three tested water purification plants around the four seasons

IIOAB JOURNAL
ISSN: 0976-3104

Algae are known as a distinct group of plant-like organisms which exist in different environments. They are photo-autotrophic cells containing chlorophyll; those cells are characterized by simple reproductive structures, and their tissue is not differentiated into true roots, stems or leaves. They vary in shape from unicellular to complex multi-cellular organisms. Some are complex in growth similar to vascular plants. Algae are found everywhere (air, soil and water) and can cause detrimental problems in water purification plants and drinking water supplies. Their growth is increased in stagnant waters with exposure to sunlight and temperatures. With regard to purification process of drinking water; Algae can cause formation of such a layer of growth on water surface and on basins walls; they can hinder filtration process through filter clogging and increase chlorine demand [11].

Types of Algae: Algal growth occurs in three substantial styles: Planktonic, Filamentous and Macrophytic. Planktonic algae are unicellular single cells floating freely in water. Excessive growth of algae is called "bloom"; when occur, it can change water color to green, brown, yellowish, red or gray. Filamentous species are called thread algae; they occur as floating green threads. They are found attached to rocks and aquatic plants. Macrophytic algal species look like vascular plants with regard to appearance [14].

Algal problems in water: Algal densities increase when nutrients increased within surface waters and when water is stagnant or has a high temperature leading to rapid growth of algae; called Algal blooms. Such thing has caused serious problems regarding water purification processes for many years. If raw water used as resources for drinking water purification process contains algal blooms, clogging of filters can occur and odors may arise in treated water. Algal blooms may cause toxins to be released in water; such toxins might be secreted by the cyanobacteria species of algae as known as blue-green algae which can cause also a bad odor to water, scum layer on the surface and may cause a risk to human health. If increased in density, Green algae may change water color to green. Therefore algal problems are mostly undesired and algae monitoring and control is very important [12, 13 and 14].

Algae Growth Conditions: Excessive algae growth will affect other forms of aquatic plant life, and can block sunlight necessary for their adequate growth. Excessive blooms of planktonic algae can also cause taste and odor problems in drinking water. Filamentous algae and macrophytic algae often form dense growths which would have a negative effect regarding fishing, swimming, and other recreational uses. Total coverage of water surface can restrict sunlight penetration and limit the production of oxygen and food items necessary for growth of aquatic habitat [14].

Changes in Phytoplankton Density: Considering the increase in total phytoplankton number was mainly attributed to diatoms which represents around 60 % of total algal counts of Nile water of Mahalet Abo Ali water purification plant, about 75 % of total algal counts of Nile water of Fewwah water purification plant and between 60 -65 % of total algal counts of Nile water of Mettobus water purification plant. Highest numbers of Diatoms were recorded in summer season indicating the effect of temperature which supports the growth of these species; similar results were given by [9 and 11]. Green algae restricted their growth to the summer season and comprise about 20% of total algal counts of raw water of the three tested sites; that agrees with Sinada et al. [15], which causes odor and taste in purified water. The high count of green algae in this season might refer to clarity of water (low turbidity) that transmit the light and the suitable temperature of the summer. Also the green algae prefer the minimal nutrients that agree with Nour Aldeen et al. [16]. Some kind of green algae can develop in summer period (high temperature) with high pH value close to that reported by Sinada et al. [15]. The alkalinity was reached the high values and so the pH that agree with NourAldeen et al. [16]. Fortunately, blue-green algae rarely grew to significant numbers during the investigated period. Blue-green algae had its maximum density in summer season with 10 – 20 % of total algal counts of raw water of Mahalet Abo Ali and Mettobus water purification plants and 6 – 12 % with regard to raw water of Fewwah water purification plant; such findings were given by [11]. Generally, the presence of high proportion of diatoms in Nile water algae indicated that, water purification plants will expect to expose frequent taste and odor problems and reduction in filter runs; however the tested treated water of the three plants had expressed a very well efficiency of the plants in algal removal process, the efficiency of the three purification plants were recorded over 90 % through out the four seasons of the year [11].

Dominant Species of Algae: There were 13 species of Diatoms had been found during the period of study as shown in [Table 3]. Findings of this study illustrated that the most dominant species of diatoms were *Syndra sp.*, *Melosira sp.*, *Diatoma sp.*, *Stephanodiscus sp.* and *Cyclotella sp.*; concerning chlorophyta, they were the second group of abundance all over the four seasons. Results of this study showed that there were 13 species of green algae which were identified and proved to be existed in raw water of the three sampling sites around the four seasons. The most dominant species were *Scendesmus sp.*, *Pediastrum sp., Actinastrium sp.* and *Kirchnella sp., and* concerning cyanophyta, their numbers were relatively lower than other groups, and results of this study indicated the existence of five species during four seasons as shown in [Table 3]. Results of this study indicated that the most dominant species were *Chrocooccus sp., Merismopodia sp.* and *Oscillatoria sp.*

IIOAB JOURNAL
ISSN: 0976-3104

Table 3: Variations in cell number (org/L) of the recorded phytoplankton species of raw water sites all over the year

Species	Mahalet Abo Ali water purification plant				Fewwah water purification plant				Mettobus water purification plant			
	winter	spring	summer	autumn	winter	spring	summer	autumn	winter	spring	summer	autumn
Diatom count	**4500**	**4900**	**6400**	**5300**	**4900**	**5700**	**8000**	**7850**	**4200**	**4850**	**7500**	**6350**
Asterionella sp.	375	420	580	564	399	538	624	590	305	365	550	530
Cyclotella sp.	455	390	450	365	357	387	670	649	450	478	752	654
Stephanodiscus sp.	480	240	654	458	368	488	825	792	320	350	558	418
Diatoma sp.	450	366	500	365	348	400	601	568	478	480	601	450
Nitzchia sp.	600	510	535	412	298	405	587	567	416	478	560	355
Melosira sp.	175	350	398	318	351	378	599	566	300	315	470	440
Syndra sp.	212	322	520	420	398	348	650	687	218	387	610	589
Cocconies sp.	165	220	400	388	427	454	501	497	388	401	580	560
Navicula sp.	420	512	687	550	569	597	487	565	208	351	690	516
Amophora sp.	365	385	485	410	378	466	509	488	165	284	517	380
Surirella sp.	223	294	330	290	354	394	635	621	388	270	400	408
Gyrosigma sp.	314	237	365	362	287	395	677	662	314	320	602	490
Fragilaria sp.	266	654	496	398	366	450	635	598	250	371	610	560
Green algae count	**1200**	**1400**	**2250**	**2200**	**1650**	**1400**	**2000**	**1700**	**1650**	**1910**	**2450**	**1730**
Crucigenia sp.	91	115	185	165	131	107	140	110	149	171	208	140
Selanastrum sp.	108	118	252	218	125	120	165	140	138	164	187	154
Scendesmus sp.	85	93	191	180	106	101	187	114	125	148	192	130
Pediastrium sp.	170	182	250	231	190	174	180	185	136	166	188	174
Tetraedron sp.	101	120	170	184	150	135	165	172	143	158	210	168
Oocyst sp.	74	85	135	143	132	91	137	124	115	124	194	121
Botryococcus sp.	77	90	156	162	114	96	176	161	104	140	200	104
Straurastrum sp.	66	84	145	162	95	90	148	143	106	120	150	103
Actinastrium sp.	80	92	167	154	126	84	190	91	110	176	151	93
Kirchnella sp.	75	81	90	92	96	76	102	80	107	152	172	141
Coelastrum sp.	45	75	94	87	87	69	94	84	90	104	130	97
Ankistrodesmus sp.	60	74	97	108	85	71	86	99	93	100	188	83
Microactinum sp.	98	105	185	174	120	90	118	87	129	107	120	115
Anthophysis sp.	70	86	133	140	93	96	112	110	105	80	160	107
Blue green count	**1100**	**900**	**2210**	**2000**	**1300**	**900**	**1450**	**650**	**1150**	**1740**	**1150**	**1720**
Chrococcus sp.	222	210	580	558	261	220	320	109	275	377	184	353
Oscillatoria sp.	280	190	531	497	269	201	291	115	254	391	299	352
Microsystis sp.	87	100	296	188	284	161	271	130	180	305	193	345
Merismopodia sp.	270	214	393	371	276	168	322	157	231	287	256	384
Ulothrix sp.	241	186	410	386	210	150	246	139	210	380	218	286

CONCLUSION

Drinking water purification plants in Kafr El-Sheikh Governorate are affected by the excessive discharge of many types of pollutants through the resources of surfaces water used as sources for drinking water purification; causing many problems with regard to the performance of the purification plants and the health of consumers. Algae might be one of the biggest problems of fresh water in Kafr El-Sheikh governorate, Egypt. This study monitored the seasonal variation of fresh and treated water phytoplankton density of three water purification plants located on Nile River in Kafr El-Sheikh governorate and showed that phytoplankton density increased in spring and summer seasons more than autumn and winter seasons; while efficiency of algal removal process exceeded 90 % for the three tested plants. The study also accentuated the most dominant species of algae which might be found in raw water of the area of study. This study recommends the continuous follow-up of counting, classification and identification of algae with regard to the resources of fresh water for early detection of problems and proper preparation in order to avoid such troubles at the technical and operational levels of purification plants to avoid the risks resulting from increased numbers of algae or the emergence of new species which may cause dangerous dilemmas to maintain the health of citizens and the integrity of purification process.

CONFLICT OF INTEREST
There is no conflict of interest.

ACKNOWLEDGEMENTS
None

FINANCIAL DISCLOSURE
None

IIOAB JOURNAL
ISSN: 0976-3104

REFERENCES

[1] Shehata, Salwa A, Ali Gamilla H, Wahba SZ. [2008] Distribution Pattern of Nile Water Algae with Reference to its Treatability in Drinking Water. Journal of Applied Sciences Research, 4(6): 722-730.
[2] Robarts RS. [1985] Hypertrophy, a consequence of development. International Journal of Environmental Studies, 12: 72-89.
[3] Reynolds CS. [1992] Eutrophication and the management of planktonic algae: what Vollenweider couldn't tell us, Eutrophication: research and application to water supply. (Eds D.W. Sutcliffe & J.G. Jones), pp 4-29. Freshwater Biological Association, Ambleside, Cumbria.
[4] Vasconcelos VM. [2001] Toxic freshwater cyanobacteria and their toxins in Portugal. In: I. Chorus, Editor, Cyanotoxins—Occurrence, Effects, Controlling factors, Springer, Berlin, 64-69.
[5] Statzner B, Bis B, Dolédec S, Usseglio-Polatera P. [2001] Perspectives for biomonitoring at large spatial scales: a unified measure for the functional composition of invertebrate communities in European running waters. Basic and Applied Ecology, 2: 73-85.
[6] Simboura N, Panayotidis P, Papathanassiou E. [2005] A synthesis of the biological quality elements for the implementation of the European Water Framework Directive in the Mediterranean ecoregion: the case of Saronikos Gulf. Ecological Indicators, 5(3): 253-266.
[7] Ekdahl EJ, Teranes JL, Wittkop CA, Stoermer EF, Reavie ED, Smol JP. [2007] Diatom assemblage response to Iroquoian and Euro-Canadian eutrophication of Crawford Lake, Ontario, Canada. J of Paleolimn., 37: 233-246.
[8] MWRI [2002] Survey of Nile system pollution sources. report no. 64, agricultural policy reform program.
[9] Kheiralla KM, Aatiga E, Elzien SM, Saud SA, Al-Imam OA. [2014] Seasonal Variation of Algae Types, Counts and Their Effect on Purified Water Quality Case Study: Al-Mogran and Burri Plants, Khartoum State, Sudan. Journal of Biodiversity and Endangered Species, 2(2): 1-4.
[10] Eaton A, Clesceri LS, RiceEW, Greenberg AE, Franson M. [2005] APHA: standard methods for the examination of water and wastewater. Centennial Edition., APHA, AWWA, WEF, Washington, DC.
[11] Shehata SA, Badr SA, Ali GH, Ghazy MM, Moawad AK, Wahba, SZ. [2009] Assessment of Nile Water Quality via Phytoplankton Changes and Toxicity Bioassay Test. Journal of Applied Sciences Research, 5(12): 2083-2095.
[12] Eutrophication of Freshwaters. [2000] Foundation of water research, Retrieved from http://www.fwr.org/waterq/frr0002.html.
[13] Cyanobacterial Toxins (Cyanotoxins) in Water. [2004] Foundation of water research, Retrieved from http://www.fwr.org/environw/frr0009.html.
[14] Algae Problems. [2016] Retrieved from https://www.lgsonic.com/blogs/algae-problems.html.
[15] Sinada FA. [1972] Ecology and productivity studies on the phytoplankton on the blue and white Nile at Khartoum. M.Sc. Khartoum University.
[16] Nour Eldeen EE, Zayed MA, Rabie KA, Moattassem ME. [1995] The seasonal variation of the Nile water quality. Journal of Environmental Science and Health A. (30): 1957-1972.

IIOAB JOURNAL

ARTICLE

AN EXTENSIBLE FRAMEWORK USING MOBILITYRPC FOR POSSIBLE DEPLOYMENT OF ACTIVE STORAGE ON TRADITIONAL STORAGE ARCHITECTURE

Prashant Desai*, Naveenkumar Jayakumar

Department of Computer Engineering, Bharati Vidyapeeth (Deemed to be University), College of Engineering, Pune, INDIA

ISSN 0976-3104

ABSTRACT

Background: *Nearly 2.5 quintillion bytes of data is generated every day and will increase ten folds by the year 2020 when Internet of Things (IoT) takes global domination in today's digital universe. Ninety percent of data generated is of "unstructured type" consisting of videos, audios, logs and sensor data, which are difficult to process. Because of the large amount of data generated and I/O intensive task being created, data transfer rate are proving to be bottlenecks in performance.* ***Methods:*** *In this research paper, we propose an active storage architecture using MobilityRPC for offloading code near to the data logic.* ***Results:*** *Both the traditional and proposed system is tested against the same synthetic workload and a reduction of 35.36 % in execution time on the proposed system is achieved.* ***Conclusions:*** *To conclude its fair say that the proposed active storage model works very well for data intensive applications where I/O bound scans are more and offloading the application is rather beneficial than the traditional method.*

.

KEY WORDS
Active storage, MobilityRPC, Application offloading, data intensive

INTRODUCTION

With the proliferation in processor speed and storage capacity, we assess Active Storage architecture as an alternative that exploits processing capabilities embedded in a disk drive to execute application. Moreover, see it has a viable and a profitable option to move computation closer to where data resides. Active Disks, which are based on the same ideology as that of Active Storage architecture of offloading computation to disk drives comes with numerous advantages such as less I/O bound scans, a higher degree of parallelism and I/O interconnect because of each disk having its own processor [1]. However, there are certain disadvantages such as system with large configuration; the disk can communicate only through the host interface leading to bottlenecks and its inability to process complex applications. Intelligent Disks (IDisks) can be seen as the successors of Active Storage with the ability to process complex applications at the cost of higher hardware resources [2]. Hadoop is a relatively new storage paradigm, which embraces the concept of Active Storage. It offers ample of advantages such as scale-out architecture, reliability, capability to store huge volumes of data and processing logic closer to data rather than the traditional way.

One of the principal techniques employed by the architectures mentioned above is offloading or migrating of application code using techniques such as code mobility, code on demand, remote evaluation and mobile agents to execute an application on the storage device. So using a mobility framework, we propose an architecture that encompasses active storage principals. The rest of the paper is divided into the following sections; a survey of Active Storage concepts and technologies, architecture of the proposed system and the technologies used implementation of the proposed system and a comparative result analysis with a traditional system.

Received: 24 April 2018
Accepted: 20 May 2018
Published: 2 June 2018

LITREATURE REVIEW

Acharya *et al.*propose a stream-based programming model that is designed to take advantage of the Active Disk architecture. Active Disk function by partitioning an application into host-occupant and disk-occupant components. The data-intensive tasks of the applications are offloaded to the disk- resident processors, while the entire task is managed and coordinated by the host-resident processors. The proposed stream-based model is for the disk-resident code also known as Disklets [1] and its communication with the host. A Disklet is language independent, take data stream as input, and produce data streams as output. It is designed to run in a sandbox with restricted privileges such as, cannot allocate memory and initiate I/O calls with the system. The Active Disks require the support of operating system at both, system and disk level, the proposed stream-based model takes advantage by simplifying resource allocation; it assigns memory in contagious blocks and provides stream buffers for communication [1]. A number of algorithms are proposed to compare and benchmark the performance of Active disk to conventional disks.

Keeton *et al.* propose an architecture for decision support systems (DSS) in which the servers use disk embedded with the microprocessor for offloading computation from the host processor. The disks are known as IDISKS [2] and are interconnected by high-speed crossbar switches for communication. It offers several advantages in terms of cost and performance over traditional server architecture. By embedding processor on the disks, computation is moved near data reducing data movement through the I/O system. The disks are interconnected to a high-speed switch thus eliminating the I/O bottleneck of the conventional disk and allowing storage devices to truly scale with data growth.

***Corresponding Author**
Email:
prashantdesai99@gmail.com
Tel.: +91- 8806507647

IIOAB JOURNAL
ISSN: 0976-3104

Anastasiadis *et al.*propose Lerna [3] an Active storage framework for executing application close to the data. The goal is to create an execution environment within the network file server that will allow the users and administrators to run executable code. In the architecture, multiple storage nodes in a server are arranged into multi-tier hierarchy assembly. The client mounts the servers as normal remote file servers, and the files and directories can be accessed with transparency. The proposed architecture uses URL encoding scheme to convert command line statements into file name strings [3]. The remote services can receive data from client machines by remote method invocation.

Ma *et al.*propose a multiview storage system (MVSS) framework [4] for offloading services to storages devices. The concept of MVSS views is similar to database views; they are generated dynamically and not stored on the storage devices. These views are presented to the users through the file system namespace. It uses block interface that supports a wide range of platforms allowing the reuse of existing file system. It introduces the concept of virtual files; it is the combination of files and certain services. Virtual disks are an abstract version of physical drives and contain the virtual files. In MVSS, new services can be dynamically created from existing ones by downloading apiece of code onto the device [4]. To add write support and cache coherency it MVSS implement a lightweight file access protocol, which is not efficient to previous implementations but simple.

Riedel *et al.*propose a system called Active Disks for exploiting the processing power of disks to run application code. They proposereal-time applications such as nearest neighbor search, data mining and edge detection [5] for benchmarking the Active Disks. The basic approach is to utilize to thedegree of parallelism offered by Active Disks, operate with minimal state information, process data while streaming from disks directly and execute arelatively small number of instructions per byte. The speedup offered by the proposed analytical model is twice of that of conventional disk drives. For sorting algorithms, data is read into the nodes from the local drives exchanged across the network according to key space distribution and sorted on the local nodes independently.

Joo *et al.*propose a virtual machine based platform IOLab, [6] for executing intelligent functions efficiently on active storage devices. It has negligible performance overhead and provides theuser with latest intelligent functions.IOLab is a user space module that is interpolated between a virtual machine and the virtual file system (VFS) of a host. IOLab intercepts I/O requests from the VM and forwards it to the intelligent functions executing in the IOLab. IOLab is implemented, as an application on the host hence does not reply on dedicated driver support. It implements the concept of active storage by running commodity block devices. The block cache stores the data blocks requested by the VM and are serviced by IOLab instead of being redirected directly to active storage. The I/O dispatcher performs operations such as data replication, migration, and prefetching in the IOLab. By defining the maximum block cache size the dispatcher rate can be varied.

Runde *et al.* propose a framework for the execution of object storage devices downloaded from the host. The framework is based on the use of virtual machines for providing processing capabilities and executing the offloaded code. The framework is built upon an open source OSD stack, with the core of the framework running the desired virtual machine [7]. The current version of the platform has a C API engine running on a Linux OSD. The VMs are sandboxed to prevent unauthorized execution of code snippets. RPC model is the central programming model for this framework. In order to implement commands on active storage remotely the OSD functions are modified with execute_function command that executes functions that exist as objects on the storage [7]. The proposed platform provides support for multiple engines allowing simultaneous execution of functions and offers high degree of parallelism.

Yin *et al.*propose architecture for active storage systems and a hybrid disk configuration for the storage system in active storage. Termed as the HcDD the disk architecture consist of dual buffer; [8] a HDD write buffer for deduplication of write requests and SSD buffer for providing parallelism to write requests. Read requests are directly processed by the SSD whereas write request to an active storage system is sent to the HDD buffer for deduplication before processing. The major contribution of this research is to reduce the number of write requests sent to the SSDs thereby reducing the number of erasure cycles, and therefore increasing the lifespan of SSDs. The proposed model integrates HDDs and SDDs to create hybrid disk architecture, before the write requests are written they deduplicated in the active storage. The processor on the disk calculates the hash value for each page in the buffer. The hash value is used as a fingerprint to lookup data on the SSD.

PROPOSED SYSTEM ARCHITECTURE

The proposed system architecture for an active storage is given below in [Fig. 1]. We use the following technologies for implementing the proposed system.

Virtualization: The data node is virtualized using a bare metal hypervisor [9] also known as the type-1 hypervisor, this type of virtualization provides complete control over the hardware, provides better scalability and performance compared to hosted hypervisors [9]. It also has the ability the host operating systems. VMware ESXi is used to host an operating system. The application migrated from the application node is offloaded and executed on this hosted operating system in contrast to the traditional system which involves moving data across the network on demand. VMware ESXi provides a host of tools for monitoring and recording the performance of operating systems.

IIOAB JOURNAL
ISSN: 0976-3104

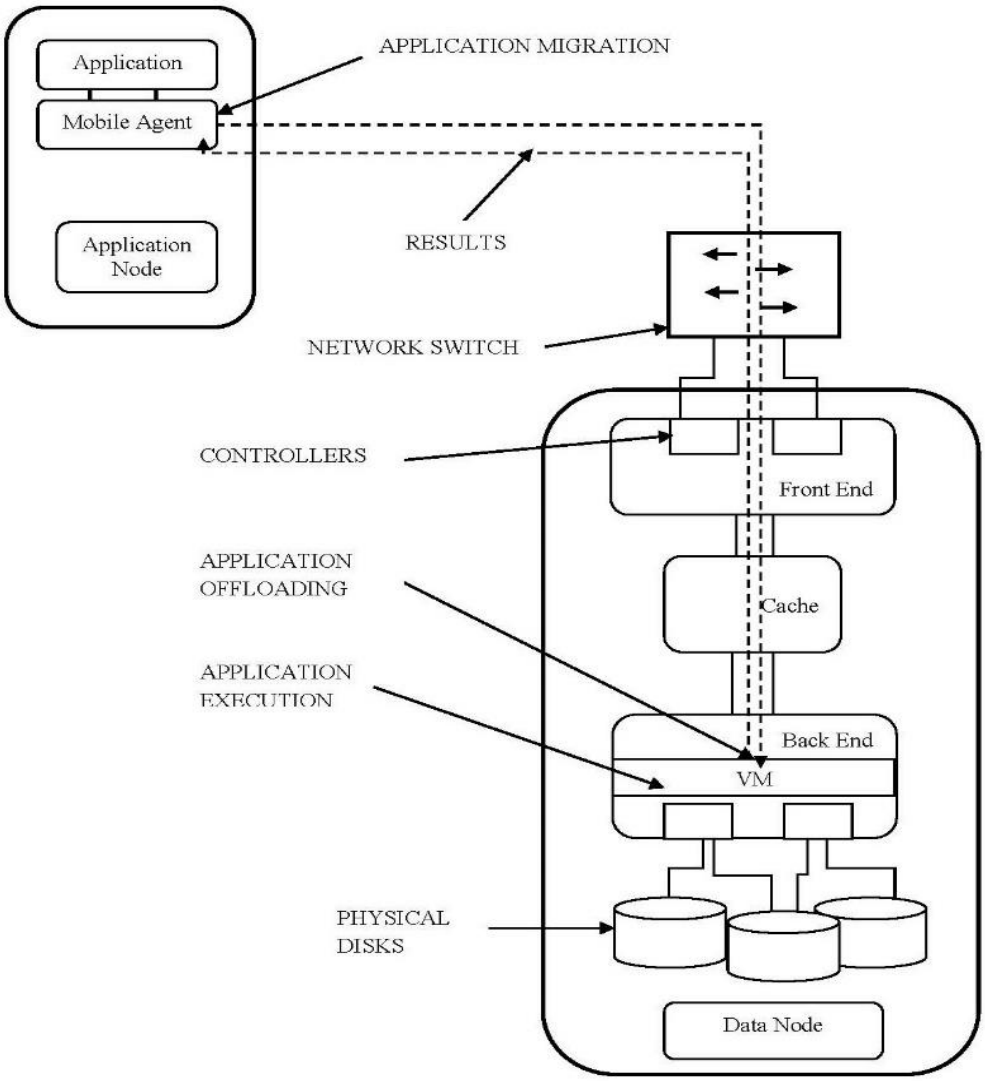


Fig. 1: Proposed system architecture.

MobilityRPC: It is a mobile agent framework used for migrating code between two nodes. It has the ability to write objects, which can move within a network and invoke arbitrary methods remotely [10]. It facilitates communication between two machines by defining a protocol, which allows class loaders to communicate and exchange bytecode. During serialization of an object all, its data is saved in a file and the objects, which it references recursively in an object graph. When the object migrates to different location and the JVM tries to deserialize it, the deserializer will try to create object instances for that class from the available data recreating the object graph on this machine. In situations where the class does not exist in the JVM, the deserializer will send request to the class loader of the application where it was originally serialized to provide for the class file. Since the class is not available on that machine, the class loader will not be able to return the class.

MobilityRPC overcomes this problem by allowing users to write their own class loaders and configure the deserializer to invoke custom class loader to provide classes needed to deserialize an object. If the classes are not present on the other machine, the custom class loader is configured to retrieve classes from anywhere we program it to. In the case of code mobility, it retrieves classes across the network. MobilityRPC defines a protocol to serialize objects instead of using Java's inbuilt serialization technique. It uses Kryo [11] a more efficient and flexible serializer, and its most important feature being it can be configured to load classes using custom class loader. Compared to other mobility frameworks that use Java RMI for communication, MobilityRPC uses custom defined RPC network protocols for exchanging messages. The most important messages are, Execution Request() - it encapsulates a serialized object wrapped in a Runnable or Callable object [12]. It sends request to MobilityRPC libraries on remote machine to deserialize the object and invoke the run() method in it. ResourceRequest() – this call allows a remote machine to request for a resource (bytecode) from the source machine [12].

IIOAB JOURNAL
ISSN: 0976-3104

IMPLEMENTATION

The execution is carried over two differently configured setups having different testbed setup ups whose performance is measured under the same synthetic workload. A traditional system is one that stimulates a normal distributed processing system, where data across the network is fetched for execution. On the other hand is the proposed system architecture, which uses the mobility RPC framework for offloading the code on the data node and carrying out the execution. The traditional system and the proposed system architecture are given below in the [Fig. 2] and [Fig. 3].

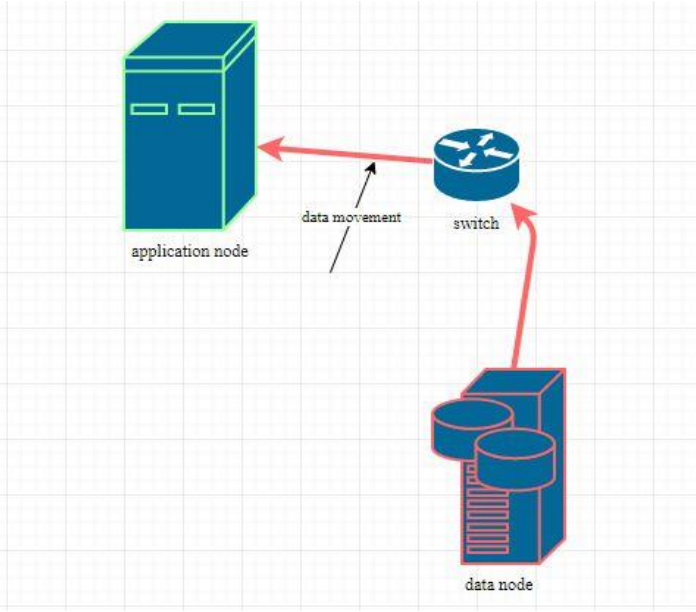


Fig. 2: Testbed1: traditional system architecture.

Testbed1 configuration: Testbed1 is configured in a normal client server configuration. The application runs on the application node and requests for data through the network from the data node. The application is written in Java, it runs through NetBeans, an IDE for running the application, and the database connectivity code is in MYSQL hence uses JDBC drivers to access the database on the storage node. It uses a TCP/IP connection specified by the IP address and port no of the data node to connect and exchange data.

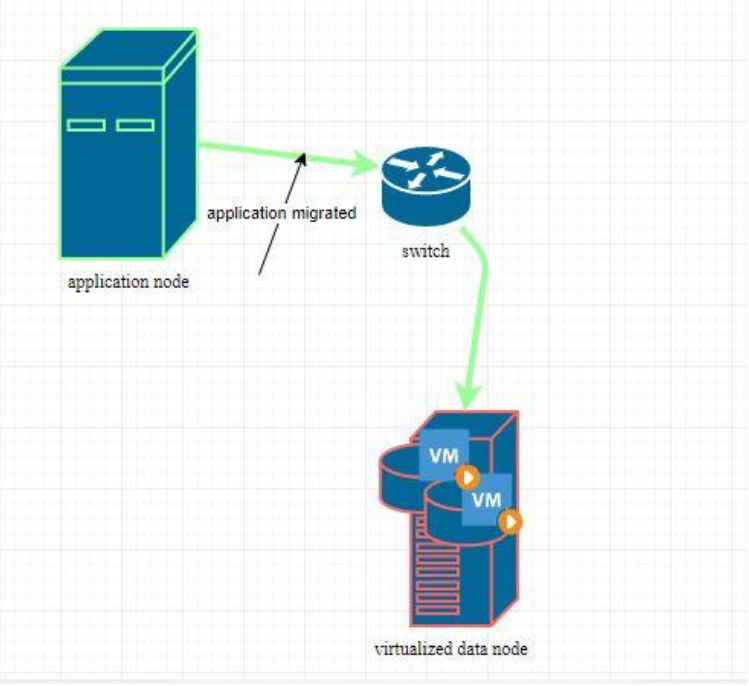


Fig. 3: Testbed2: proposed system architecture.

IIOAB JOURNAL
ISSN: 0976-3104

Testbed2 configuration: Testbed2 configuration is similar to that of Testbed1. Here the MobilityRPC framework runs on the application node inside the NetBeans IDE. The MobilityRPC standalone server is running on the data node to receive incoming mobile agents from the application node. The mobile agent on the application node carries the application code to the data node hence MYSQL is able to connect to the database using localhost cause no network is involved and executes on the data node. VMware ESXI 5.5 hypervisor is used to for virtualization on the data node and provide real time performance analysis.

For the stimulation on the traditional system, a sample MySQL employee database is used as the workload. The application is designed to execute query, which requires connecting to the database located on the data node and fetch data. This causes data to be moved across the network affecting the execution time of the application. The query is designed to be data intensive in nature and have high I/O bound requests. For the proposed traditional system, the same application executed except it is embedded on to the MobilityRPC framework. The data intensive part of the application code is migrated to the data node for execution hence stimulating the concept of near data processing or active storage. By moving the application closer to the database, it does have to open remote connections to the database and can directly process it in-situ speeding up the execution time of the application compared to that of the traditional system.

RESULTS

The SQL query executed on the employee database for both the architectures is SELECT * from employees, salaries WHERE emplyoees.emp_no=salaries.emp_no ORDER BY salary DESC, last_name ASC, hire_date DESC. The LIMIT function is used to limit the number no records retrieved. What makes this query data intensive is the ORDER BY clause, which demands the records, returned to be sorted as per the requirements. Since sorting is data intensive and the returned columns are individually sorted resulting in the entire database to be traversed. [Table 1] below gives, us a comparative analysis of the execution time required for the query on traditional system and the proposed system architecture.

Table 1: Execution times for traditional and proposed system architecture

Number of records retrieved	Execution time (traditional architecture)(sec)	Execution time (Proposed Architecture)(sec)
150000	97	98
200000	161	108
250000	205	124
300000	244	144
350000	278	183
400000	345	199
450000	372	235
500000	638	273

From the data in [Table 1], it can be easily concurred that the proposed system architecture is better for executing data intensive tasks and is the way forward. Initially when the number of records to be retrieved are less the traditional system is better because of the overhead associated with MobilityRPC, however as the size of data increases this overhead is compensated. On calculating the average from the above table, a significant 35.36 % reduction in execution time is achieved by implementing the proposed system architecture. The graphical representation of the table is given below in the [Fig. 4]

From the above results, we can conclude that the proposed framework can be integrated in MySQL as an additional feature for executing data intensive tasks such as sorting and traversing through database located on remote machines. It can find application in any database that offers remote database connectivity.

CONCLUSION

After comparing the results of the simulation on the traditional system and the proposed system architecture a 35.36% reduction in execution time is observed which hence proves that the proposed model is more efficient for executing data intensive tasks than the traditional system. Although not applicable in all scenarios, it strengthens the argument of offloading compute logic near the data to decrease the execution time instead of accessing data across the network which increases execution time in case of high I/O bound tasks. The proposed framework can be integrated into data intensive computing for efficiently executing distributed computing similar to Hadoop's HDFS for data processing. Other areas of research that are of interest for this framework would be data mining, where the application is offloaded to the data warehouse and load balancing. Creating a dynamic framework for migrating compute intensive

ISSN: 0976-3104

and data intensive tasks on demand and deploying the framework on advance hardware such as SSD's in RAID configuration remains the future scope of the research.

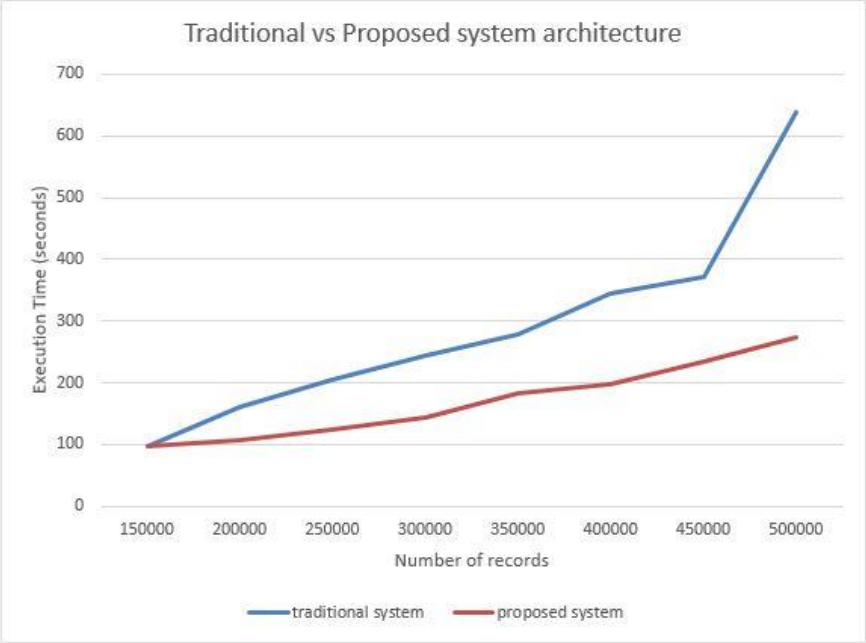


Fig. 4: traditional system v/s proposed system.

CONFLICT OF INTEREST
There is no conflict of interest.

ACKNOWLEDGEMENTS
None

FINANCIAL DISCLOSURE
None

REFERENCES

[1] Acharya A, Uysal M, Saltz J. [1998] Active disks: programming model, algorithms and evaluation, *SIGPLAN Not.*, 33(11): 81–91.
[2] Keeton K, Patterson DA, Hellerstein JM. [1998] A case for intelligent disks (IDISKs), *ACM SIGMOD Rec.* 27(3): 42–52
[3] SV Anastasiadis, RG Wickremesinghe, JS Chase.[2005] Lerna: An active storage framework for flexible data access and management, *Proc. IEEE Int. Symp. High Perform. Distrib Comput*, pp. 176–187.
[4] Ma X, Reddy ALM [2003] MVSS: An Active Storage Architecture, *IEEE Trans. Parallel Distrib. Syst.*, 14(10): 993–1005.
[5] Riedel E, Gibson GA, Faloutsos C. [1998]Active Storage for Large-Scale Data Mining and Multimedia," *Proceedings*, no. February, pp. 62–73.
[6] Joo Y, Ryu J, Park S, Shin H, Shin KG. [2014] Rapid prototyping and evaluation of intelligence functions of active storage devices, *IEEE Trans. Comput.*, 63(9):2356–2368
[7] Runde MT, Stevens WG, Wortman PA, Chandy JA. [2012] An active storage framework for object storage devices, *IEEE Symp. Mass Storage Syst. Technol.*,
[8] Yin S *et al.*[2015] HcDD: The hybrid combination of disk drives in active storage systems, *Proc. - 2015 IEEE 17th Int. Conf. High Perform. Comput. Commun. 2015 IEEE 7th Int. Symp. Cybersp. Saf. Secur. 2015 IEEE 12th Int. Conf. Embed. Softw. Syst. HPCC-CSS-ICESS 2015*, pp. 678–683.
[9] Bare metal hypervisor, [available] https://searchservervirtualization.techtarget.com/definition/bare-metal-hypervisor [20-April-2018].
[10] Mobility-RPC, [Available] https://github.com/npgall/mobility-rpc [20-April-2018].
[11] Mobility-RPC technoligies used, [Available] https://github.com/npgall/mobility-rpc/blob/master/documentation/TechnologiesUsed.md [20-April-2018].
[12] Mobility-RPC Protocol Messages [Available] https://github.com/npgall/mobility-rpc/blob/master/documentation/ProtocolMessageStructures.md [20-April-2018].