Review-based product automated recommendation system in E-commerce using improved Frequent Pattern Mining and Artificial Intelligence

Bhrantav Vora

Assit. Prof. Dept. of MCA

LDRP Institute of Technology and Research

Kh – 5, Sector 15, Near Gandhinagar ITI, Gandhinagar

Dr. N J Patel

Professor, Dept. of MCA

Acharya Motibhai Patel Institute of Computer Studies , Ganpat University, Kherva

Dr. N N Jani

Dean, Dept. of Computer Science, KSV University, Gandhinagar

ABSTRACT- E commerce recommendation system frameworks is defined as data filtering techniques used for forecasting product ratings on the basis of user likes and items. Online E commerce business has provided correlation amongst temporal evaluation and consuming behaviour while processing large amount of information. The main limitation arising in E-commerce is restricting the predictive analytics of recommendation systems in mobile commerce. In order to solve this issue, this research provides an innovative approach comprising frequent pattern algorithm and combined Cuckoo-genetic search algorithm to obtain better fitness parameter in E- commerce recommendation system. Frequent pattern algorithm is used to compare the reviews obtained amongst two users having similar ratings on products through similarity indices. In addition, Hybrid Cuckoo Search and Genetic Algorithm is used to obtain the optimal solution from the user reviews. Subsequently, WEKA and R studio is used to gather information regarding reviews on various mobile brands and their price list. Various experiments have been conducted on mobile E-commerce environment and results obtained shows that traditional system is inferior compared to proposed system in terms of simplicity and accuracy. These realizations can further improve recommendation systems, and endorse the sustainable growth in E-commerce business.

KEYWORDS-*Mobile e-commerce; WEKA; R-studio; hybrid cuckoo-search algorithm; frequent pattern mining; population.*

1. INTRODUCTION

The term E-commerce or e-business refers to the business transactions among companies, or the business transaction between companies and their customers which will be entirely or partially conducted over the internet or similar computer networks [1]. The business entirely depends on the advantages of information. Gaining of more customers, relishing, and retaining existing customers, and the appropriate prediction of the buyer behavior will further enhance the availability of products and services and hence, the profit [2]. The rapid emergence of the internet and the information technology, e-commerce based on the virtual economy has been widely attracted, and it has gradually developed into the backbone of the emerging industry.

Recommender systems were generally defined as a model utilized for recommending suitable products or services to specific users by predicting the interest of each user based on the information's available in regards to the interaction between the items and users [3]. Also, the development of recommender system is found to reduce the overload of data collected by retrieving the information or services from a large enormous data present, thereby providing personalized services. The significant role of a recommender system is the ability to determine the preferences and interests of users by considering the behavior of the user in generating personalized recommendations [4]. The recently employed recommendation techniques were given by collaborative filtering (CF), content-based (CB) and knowledge-based (KB) techniques [5], [6] and [7]. However, it was observed that the

available recommendation approaches were found to have numerous advantages and limitations such as CF has sparseness, scalability, and cold-start problems [8], while CB has overspecialized recommendations [9]. Therefore, to overcome these shortcomings by developing advanced recommendation approaches such as social network-based recommender systems, fuzzy recommender systems, context awareness-based recommendation system were developed.

In addition, the recent developments of recommendation approaches and techniques has led to rapid implementation of recommender systems for real-world system applications. The recommender systems were found to be utilized for numerous applications such as movies, music, television programs, books, documents, websites, conferences, tourism scenic spots and learning materials, and involve the areas of e-commerce, e-learning, e-library, e-government, and e-business services [10]. Therefore, the recommender system applications have been categorized into distinct domains such as e-government, e-business, e-commerce/ e-shopping, e-library, e-learning, e-tourism, e-resource services, and e-group activities. Deep personalization builds a customer relationship over time, leveraging the history developed to provide better recommendations. This deep personalization uses user-user interaction, attribute-based systems along with a learning module to identify user interests. The rise in popularity of the review aggregating websites has led to an influx of the data on customer's preferences [11]. The huge repositories of the user written reviews create opportunities for new type of recommendation system that could control the content embedded in the text.

2. RELATED WORK

Throughout the years, recommender frameworks have been contemplated broadly and are separated into various categories. The review on various effective techniques such as probabilistic model, demographic based recommendation, frequent pattern-based algorithm,

A new probabilistic model called FLAME has been proposed by author [13] by the combination of advantages obtained from Aspect Based Opinion Mining and Collaborative filtering in order to evaluate the products and service offered in E-commerce websites. The author mainly concentrates on the problems arising in estimation of personalized sentiment polarities on the basis on the basis of different aspect of the item set. The study shows that experiments has been conducted on two online review datasets and result in terms of RMSE= 0.980 outperforms other existing techniques on the basis of identification and aspect rating prediction. Furthermore, a new collaborative recommendation system which generates recommendations for item set for users has been proposed by author [14]. This algorithm comprises with the advantages of collaborative filtering, dynamic content-based filtering, opinion mining and association rules. The author conducted experiments on live online dataset using precision evaluation metric and result demonstrates that recommendation produced by the proposed method outperforms other existing benchmark recommendations methods with a precision value of 76% for 15 recommendations. An effective automated collaborative filtering approach has been proposed by author [12] to combine the results of recommender system. This new algorithm comprises with a evolutionary approach called as interline in order to automate choice of techniques used by combining results of various recommendation techniques. The study shows that experiments has been conducted with a set of dataset obtained from MovieLens and different approaches of collaborative filtering and observed that an improvement varies from 3.6% to 118.9%, which increases the accuracy of the generated recommendations.

A new demographic-based product recommendation system has been proposed by author [15] by relating a new concept "We Know What You Want to Buy". MErchanT Intelligence Recommender System (METIS) technique is designed on the basis of microblogging service platform and is used to detect the number of users, purchase users through set of microblogs and online reviews. The study shows that author considered an input dataset comprising of 5 million users with 1.7 billion tweets and observed that METIS technique is feasible with very modest throughput 2.16 tweets per second. Furthermore, the process involved in exploring the demographic information has been proposed by author [16]. This research was mainly concerned on the product recommendation and its influence on the users and sales. The author has considered some preliminary concepts such as purchase-intent tweet, demographic products, and user profile in order to increase the effectiveness of the system. The study shows that experiments has been conducted data crawled large scale microblog and on Sina Weibo platform and observed that the proposed system is effective in determining the accurate recommendations with improved matching user's preference. Author has proposed a new bootstrapping approach [17] in order to extract the product adopters from a set of review text and differentiating them on the basis of demographic features. Author used graph technique to update the user number, related product distribution in the matrix factorization feature set, which is further used for product recommendation. From this study, the results on large dataset shows that there is an 6.9% increase in the performance with value 0.097 and shows that proposed method is more prominent with limited number of training data set.

In order to solve the classical problem arising in shopping recommendations, the author [18] has proposed a new type of recommendation system by comprising clustered feature sets. The shopping recommendations has been developed on the basis of global and local influences. From this study, it has been observed that easy adaptation can be obtained by considering frequent behaviour in recommendations generation. Further by using this algorithm, the application of the user profile can be extended determining salesman's public relations. A novel based Apriori algorithm has been proposed by author [19] in order to increase the value of electronic commerce recommendation system. Apriori algorithm is used to determine frequent item set by expanding the number of item set and processing through association rules to satisfy the minimum confidence level. From this study, the results show that time taken by the FP growth

algorithm and Apriori algorithm for 100 data size is 11.011 s, which is less compared to apriori algorithm and advantageous in processing large data set. Furthermore, recommender for selling associated items and to increase the promotion of sales, author has proposed frequent pattern-based algorithm [20]. In this research, author has given important concern for selling associated items and to forecast the frequent pattern on the basis of customer interest of item. From this study, the results obtained in terms of precision-32.11, F-measure-39.96 shows that the proposed system is accurate compared to other existing techniques.

Author has proposed a new artificial intelligence-based algorithm [21] to improve business performance in china's E-commerce. Artificial intelligence approaches are found to be beneficial in E-commerce environment for developing B2B and B2C E-commerce systems. The author developed a structure for Taobao's Database management system by combining AI and BPM. The study demonstrates that by these combination, there is an improvement in the performance of the system and further advancement of AI leads to implementation of better product recommendation algorithm in E-commerce. Furthermore, author has proposed a new paradigm Business Intelligent Framework [22] in order to solve the issues regarding Recommender Systems. The author has considered a high-level model steps for input collection, formatting data, application of machine learning component and data warehouse component. From this study, it has been observed that proposed framework make use of OLAP reporting tools and BD to monitor the performance of E-recommendation system. Further the features provide highly trusted and context-aware set for future recommendations.

The above study states that by developing a algorithm comprising frequent pattern mining and artificial intelligence shows high precision, processing speed and better intelligence technique for product recommendation in E-commerce. Further study and improvement in terms of hardware tools and techniques can improve the performance of the system.

3. RESEARCH METHODOLOGY

In this paper, the input is obtained from the amazon file which comprises with customer reviews on different mobile manufactures such as Samsung, Nokia, Jethro, e passion, Apple, BlackBerry, BLU, CNPGD, Quality Technology Industrial Co., ltd., HTC, Huawei, indigi, LG, LG Electronics MobileComm USA, LGIC, LifeProof, Sudroid, Posh Mobile in terms of parameters such as price, rating, review votes. The dataset obtained is tabulated in CSV excel sheet and further processed as an input to Waikato Environment for Knowledge Analysis (WEKA). Further, the data is processed to R studio in order to process large set of data. Frequent pattern algorithm is used for analysing and clustering the input data followed by efficient cuckoo search- genetic algorithm for optimization and predicting better recommendation system.

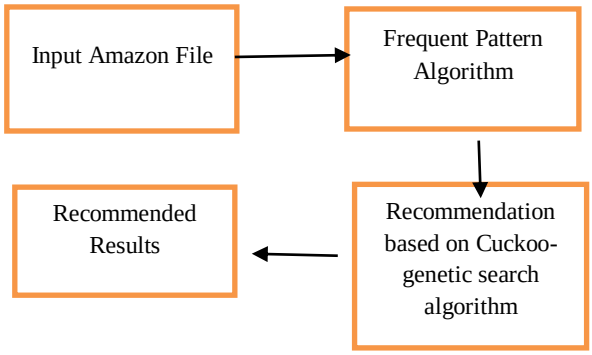

Fig 1. Block diagram of the proposed system

The block diagram of the proposed method is shown in Fig. 1. The steps involved from processing input data to recommended results are as follws,

3.1. INPUT AMAZON FILE

The input amazon file comprises with 13841 number of customer reviews on the multiple attributes. The first attribute is based on various brand names such as Samsung, Nokia, Jethro, e passion, Apple, BlackBerry, BLU, CNPGD, Quality Technology Industrial Co., ltd., HTC, Huawei, indigi, LG, LG Electronics MobileComm USA, LGIC, LifeProof, Sudroid, Posh Mobile. The second attribute comprises with the price list of their products, third with third with customer ratings and fourth with customer reviews. These attributes is tabulated in CSV excel sheet and given as input to WEKA analysis tool.

WEKA is a popular tool used for resolving real world data mining problems. WEKA comprises with various machine learning algorithms, java program is used for analysing and testing is done in almost all the platforms. The input amazon file is converted into ARFF (Attribute Relation File Form) format in order to process the data and has been analysed using frequent pattern algorithm. The limitations of WEKA is that large set of data cannot be processed and problem of getting struck in GUI version. R is an open source software program used to analyse large set of data with limited amount of time. R is defined as an integrated tool comprises with a set of software's for data manipulation, calculation and graphical facilities for analysis of data. The large data object functions such as price, rating, review votes is processed in R studio platform and data obtained is transferred to frequent pattern algorithm. Further, frequent pattern mining algorithm is used to group the data and categorize them on the basis of their features set. In this research, feature set comprises with mobile brand name, price, rating and the ratings given by the customers.

3.2. FREQUENT PATTERN MINING ALGORITHM (FP ALGORITHM)

FP algorithm is found to be an effective and efficient technique for data mining in terms of frequent patterns and growth of pattern fragment. This algorithm make use of divide and conquer technique and FP tree for input data clustering. The compressed dataset obtained from the R-model is divided into a number of conditional data set modules in which each set is associated with individual frequent pattern and it is mined distinctly. By doing so, the search cost for recursive pattern can be reduced and increases the probability for selecting better recommendation system. Further in order to handle large database, the database is partitioned to smaller database called projected database and FP tree is constructed for the set of smaller database. The algorithm developed for FP tree are as follows,

FP- tree construction algorithm
Input: A comparison database DB and a least support threshold?.
Output: Frequent patterns complete set obtained?
FP_check (Subset, σ)
λ= x U σ
Generate (λ with support = x. support)
construct λ's conditional base pattern and λ's conditional FP-subset subset $_{\lambda}$
if subset $_{\lambda} \neq Ø$
then call FP_check(subset $_{\lambda}$, λ)
Initially call:
FP_check(subset, null)

The FP tree construction algorithm comprises with two scans of database construction. The collection of frequent items is done in first scan and followed by FP tree construction in second scan. From the algorithm, the amazon database in categorized on the basis of brand name, price and customer reviews. Further the data is processed to hybrid cuckoo search and genetic algorithm to obtain optimized objective function weigh factor.

3.3. HYBRID CUCKOO SEARCH AND GENETIC ALGORITHM (HCSGA)

The advantage of cuckoo search and genetic algorithm is used to find the better optimal solution for recommendation system.

The pseudocode for the proposed HCSGA system are as follows,

Begin,

Selection of objective function FP (x);

Step 1: Initialization: Counter set to b=1; randomly initialize population.

(Initialize N_b number of host branches in the frequent pattern tree randomly such that each branch will have sub branch, which corresponds to a particular solution for a given problem);

Step 2: Evaluation of Fitness: Evaluate the fitness parameter FP (x);

While (n < maximum generation) or (end criteria); / **Generate New Population**/

Select new population by considering genetic operators (selection, crossover, mutation)

Calculate fitness (best individual population is considered)

Initiate new solution (Y_{new})

Calculate weight factor/fitness FY_{new} ;

Randomly select a solution (Y_j) among from N_{bnew} and determine its fitness (F_j);

If ($FY_{new} < F_j$) then

Replace j with new solution;

End if

Extract and store the best solution;

n = n+1;

Step 3: end while

Step 4: Evaluate entire generation set, extract best weight solution amongst current best solution stored.

End

The main problem arising during optimization is to handle number of constraints. By proposing efficient HCSGA algorithm, the efficient fitness weight factor is obtained. Initialization is done by setting random counter value as 1, the fitness parameter is evaluated through number of population generation and by considering the genetic operators. The best fitness parameter is evaluated by comparing present fitness value with the existing fitness value.

4. RESULT AND DISCUSSION

In order to evaluate the performance of the system, the input amazon file is considered comprises with 13841 review components of various mobile manufactures such as Samsung, apple which are divided into four attributes namely brand name, price, rating, review rates. The result obtained from the proposed algorithm are shown below,

```
19:26:39: RScriptExecutor$39508542|Creating rules data frame
19:26:39: RScriptExecutor$39508542|Converting rules to binary
19:26:39: RScriptExecutor$39508542|Running cuckoo search
19:26:39: RScriptExecutor$39508542|lu=2.37673560366966
19:26:39: RScriptExecutor$39508542|Running genalgo
19:26:39: RScriptExecutor$39508542|Training neural net with layers = 321
19:26:39: RScriptExecutor$39508542|Computing neural net accuracy
19:26:41: RScriptExecutor$39508542|bestthreshold:bestaccuracy = 0.39:0.476190476190476
19:26:41: RScriptExecutor$39508542|Training neural net with layers = 312
19:26:42: RScriptExecutor$39508542|Computing neural net accuracy
19:26:43: RScriptExecutor$39508542|bestthreshold:bestaccuracy = 0.42:0.523809523809524
19:26:43: RScriptExecutor$39508542|Training neural net with layers = 321
19:26:43: RScriptExecutor$39508542|Computing neural net accuracy
19:26:45: RScriptExecutor$39508542|bestthreshold:bestaccuracy = 0.39:0.523809523809524
19:26:45: RScriptExecutor$39508542|Training neural net with layers = 312
```

Fig. 3. Data acquisistion in R studio

19:26:39: RScriptExecutor$39508542|Training neural net with layers = 321
19:26:39: RScriptExecutor$39508542|Computing neural net accuracy
19:26:41: RScriptExecutor$39508542|bestthreshold:bestaccuracy = 0.39:0.476190476190476
19:26:41: RScriptExecutor$39508542|Training neural net with layers = 312
19:26:42: RScriptExecutor$39508542|Computing neural net accuracy
19:26:43: RScriptExecutor$39508542|bestthreshold:bestaccuracy = 0.42:0.523809523809524
19:26:43: RScriptExecutor$39508542|Training neural net with layers = 321
19:26:43: RScriptExecutor$39508542|Computing neural net accuracy
19:26:45: RScriptExecutor$39508542|bestthreshold:bestaccuracy = 0.39:0.523809523809524
19:26:45: RScriptExecutor$39508542|Training neural net with layers = 312
19:26:46: RScriptExecutor$39508542|Computing neural net accuracy
19:26:47: RScriptExecutor$39508542|bestthreshold:bestaccuracy = 0.37:0.476190476190476
19:26:47: RScriptExecutor$39508542|u=2.36594690568745
19:26:47: RScriptExecutor$39508542|Running genalgo
19:26:47: RScriptExecutor$39508542|Training neural net with layers = 321
19:26:48: RScriptExecutor$39508542|Computing neural net accuracy

Fig. 4. Accuracy data in R framework

Fig.3 and Fig. 4 shows the results obtained from R-studio framework in terms of training neural network layers threshold best accuracy. Individual reviews obtained from the customers in terms of mobile is made to undergo fitness parameter check through proposed algorithm in order to predict accurate ratings of the product.

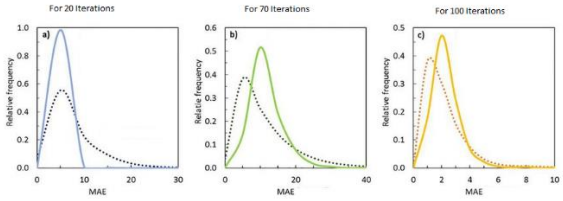

Fig 5.Plot of MAE vs Related Frequency

Individual review component obtained is processed through HCSGA algorithm in order to calculate per user item coverage, diversity and accuracy. From Fig. 5 it has been observed that by comprising 20 iterations, 70 iterations and 100 iterations, the results stabilize with increase in number of iterations with less mean absolute error. The coloured line from the graph shows the results obtained through HCSGA algorithm.

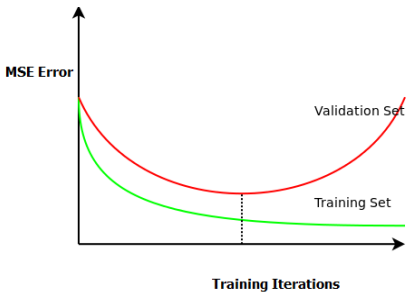

Fig 6.Plot of MAE vs Related Frequency

Fig. 6. shows the graph of MSE vs number of iteration's obtained in order to calculate mean squared error. Validation set is used as a part of training set and by proper tuning, accuracy of the system is increased with less error. From the graph, it has been shown that MSE decreases with increase in the number of iterations and leads to optimal prediction in e-commerce recommendation system.

From the experimental study, it has been observed that for individual sample an average accuracy rate of 40% and overall sample accuracy of 91.11% is obtained from the proposed mobile recommendation system.

CONCLUSION

This research provides effective mobile E-commerce recommendation system method comprising FP- tree construction algorithm and HCSGA. It provides customers with more appropriate information at less time constraint and help the customer choosing better efficient products through online E-commerce. The literature on E-commerce recommendation system provides the requirement of customers, limitation existing in E-commerce system and further provides information about the requirement of the present research through number of stages. The recommendation system will not only enhance the satisfaction in customers, but also influence the conversion rate of goods. Thus recommendation in E-commerce helps in selecting more suitable products and increases sale rates.

REFERENCES

1. Turban, E., King, D., Lee, J. K., Liang, T. P., & Turban, D. C. (2015). *Electronic commerce: A managerial and social networks perspective*. Springer.
2. Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.
3. Lu, J., Wu, D., Mao, M., Wang, W., & Zhang, G. (2015). Recommender system application developments: a survey. *Decision Support Systems*, *74*, 12-32.
4. Omondi, A. O., & Mbugua, A. W. (2017). An Application of association rule learning in recommender systems for e-Commerce and its effect on marketing.
5. Contratres, F. G., Alves-Souza, S. N., Filgueiras, L. V. L., & DeSouza, L. S. (2018, March). Sentiment Analysis of Social Network Data for Cold-Start Relief in Recommender Systems. In *World Conference on Information Systems and Technologies* (pp. 122-132). Springer, Cham.
6. HamlAbadi, K. G., Saghiri, A. M., Vahdati, M., TakhtFooladi, M. D., & Meybodi, M. R. (2017). A Framework for Cognitive Recommender Systems in the Internet of Things (IoT).
7. Zhang, Z., Lin, H., Liu, K., Wu, D., Zhang, G., & Lu, J. (2013). A hybrid fuzzy-based personalized recommender system for telecom products/services. *Information Sciences*, *235*, 117-129. [10] J. Lu, Q. Lu, J., Shambour, Q., Xu, Y., Lin, Q., & Zhang, G. (2013). a web-based personalized business partner recommendation system using fuzzy semantic techniques. *Computational Intelligence*, *29*(1), 37-69. [11] Adomavicius, G., & Tuzhilin, A. (2015). Context-aware recommender systems. In *Recommender systems handbook*(pp. 191-226). Springer, Boston, MA.

8. Masthoff, J. (2011). Group recommender systems: Combining individual models. In *Recommender systems handbook* (pp. 677-702). Springer, Boston, MA.
9. Park, D. H., Kim, H. K., Choi, I. Y., & Kim, J. K. (2012). A literature review and classification of recommender systems research. *Expert Systems with Applications*, *39*(11), 10059-10072.
10. Wu, D., Zhang, G., & Lu, J. (2015). A fuzzy preference tree-based recommender system for personalized business-to-business e-services. *IEEE Transactions on Fuzzy Systems*, *23*(1), 29-43.
11. Wang, J. (2013). *Session aware recommender system in e-commerce* (Doctoral dissertation, University of California, Santa Cruz). da Silva, E. Q., Camilo-Junior, C. G., Pascoal, L. M. L., & Rosa, T. C. (2016). An evolutionary approach for combining results of recommender systems techniques based on collaborative filtering. *Expert Systems with Applications*, *53*, 204-218.
12. Wu, Y., & Ester, M. (2015, February). Flame: A probabilistic model combining aspect based opinion mining and collaborative filtering. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining* (pp. 199-208). ACM.
13. Tewari, A. S., & Barman, A. G. (2017). Collaborative Recommendation System Using Dynamic Content based Filtering, Association Rule Mining and Opinion Mining. *International Journal of Intelligent Engineering and Systems*, *10*(5), 57-66.
14. Zhao, X. W., Guo, Y., He, Y., Jiang, H., Wu, Y., & Li, X. (2014, August). We know what you want to buy: a demographic-based system for product recommendation on microblogs. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1935-1944). ACM.
15. Zhao, W. X., Li, S., He, Y., Wang, L., Wen, J. R., & Li, X. (2016). Exploring demographic information in social media for product recommendation. *Knowledge and Information Systems*, *49*(1), 61-89.
16. Zhao, W. X., Wang, J., He, Y., Wen, J. R., Chang, E. Y., & Li, X. (2016). Mining product adopter information from online reviews for improving product recommendation. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *10*(3), 29.
17. Toma, A., Constantinescu, R., & Nastase, F. (2009). Recommendation system based on the clustering of frequent sets. *Wseas transactions on information science and applications*, *6*(5), 715-724.
18. Zhu, Q., Lu, H., & Xu, X. (2016). Research and Application of E-Commerce Recommendation System Based on Association Rules Algorithm. *Journal of Residuals Science & Technology*, *13*(5)
19. Cho, Y. S., & Moon, S. C. (2016). Frequent Pattern to Promote Sale for Selling Associated Items for Recommender in e-Commerce. *Indian Journal of Science and Technology*, *9*(38).
20. TAPSOBA, L., & XIAO, Z. (2017). Analysis of AI contribution to improving BPM of E-Commerce in China: examining the case of Taobao.
21. Venkatraman, S. (2017, November). A Proposed Business Intelligent Framework for Recommender Systems. In *Informatics* (Vol. 4, No. 4, p. 40). Multidisciplinary Digital Publishing Institute