IJCTA, 9(27), 2016, pp. 551-559 ISSN: 0974-5572

An Improved Candidate Transaction PageRank Algorithm using Classification Techniques for Web Search Engines

Srinaganya G.[1] and J.G.R. Sathiaseelan[2]

ABSTRACT

Improved Candidate Transaction PageRank Algorithm is a technique for index the web pages with more relevant and accurate information within a tiny time limit for the given user query in web search engines. This paper compares some classification algorithms for T1014D100K, Sina Weibo and Twitter datasets. The proposed algorithm is evaluated with Apriori, KNN, SVM and AdaBoost algorithms, when it is applied to the original dataset the relevancy has 99.77% with 99.22% of Accuracy and the limit of the Time is 82.15%. It indicates that the SVM and AdaBoost classifiers have better accuracy and relevancy in low time consumption. It leads to the conclusion that SVM and AdaBoost is good classifiers among these classification methods with 99.99% using Improved Candidate Transaction PageRank Algorithm.

Keywords: Search Engine, PageRank, HITS, Improved Candidate Transaction PageRank Algorithm, AdaBoost, SVM.

1. INTRODUCTION

A search engine is a software program designed to find information stored on the World Wide Web. According to the user query, the search engine retrieves a list of references that fulfils the user query [1]. So it is able to provide desired order of web pages. The search result may be a combination of web pages, images, videos and other types of files. The web pages should contain relevant information, to the searcher about the searched keyword or subject. Search engines use regularly updated indexes to operate quickly and efficiently [2]. The centralized search system is based on conventional database system. This system has its own data collecting mechanism and all the data are stored and indexed in a database system [3]. The distributed search system distributes information retrieval system, which has not its own actual database [4]. When receive a query from a user, it will instantly obtain the records through the search interfaces that is provided by sub-databases [5]. The development of the Internet and the popularity of World Wide Web makes the web page ranking system has drawn significant attention [6]. Many web search engines have been still trying to provide completely relevant answers to the general subject of queries [7].

The basic approach of PageRank is that, a document is considered by how many important in-links it have, but those in-links do not count similarly [8]. Jon Klein berg came up with his own solution to the web search problem and developed an algorithm that made use of the link structure of the web in order to discover and rank pages relevant for the particular topic [9]. Hyperlink Induced Topic Search (HITS) is now part of the Ask search engine [10]. HITS algorithm is in the same spirit as PageRank. They both make use of the link structure of the web graph in order to decide the relevance of the pages [11]. The difference is that unlike the PageRank algorithm, HITS only operates on a small sub-graph from the web graph.

2. RELATED WORK

The users of World Wide Web are increasing day to day life. The users need the relevant, accurate information from a short time duration from WWW. To rectify this problem many researchers develop so many ranking algorithms from the basic

1 Department of Computer Science, Bishop Heber College, Tiruchirappalli, Tamilnadu, India

2 Department of Computer Science, Bishop Heber College, Tiruchirappalli, Tamilnadu, India

PageRank and HITS algorithms. Some of the ranking algorithms are discussed here that are related to the proposed work. A novel approach is presented by Marc Clasesen et al. [12] to learn binary classifiers when only positive and unlabeled instances are available. Using an ensemble of SVM models trained on bootstrap resamples of the training data for increased robustness against label noise. The comparison has done with state of the art approaches in simulations using multiple public benchmark datasets. In [13], they explained machine learning and natural language processing techniques used and applications of sentiment analysis. Also they presented the open issues along with a summary table of a hundred and sixty one articles. This work was done for the social media and micro blogs using machine learning method.

In [14], they introduced a most important technique that are used for text mining tools are currently being used and type of problems they are typically applied for. The performance of two unsupervised approaches are evaluated by Noraini Seman et al. [15] was Maximum Marginal Relevance (MMR) and Concept-based global optimization framework. Automatic summarization is very useful techniques that can help the users browse a large amount of data. This paper proposed some improved methods by blending each unsupervised approach at sentence level. In [16], they showed that AdaBoost, KNN and SVM-RBF have exhibited over fitting the results when they were applied to the original dataset. There is non-over fitting occurred for the random sampling dataset where SVM-RBF had the highest accuracy for training and testing dataset.

3. PROBLEM DEFINITION

The fame web sites do not guarantee that the desired information to the searcher with the relevant factor. In Internet, available data is huge and the algorithm is not fast enough [17, 18]. It should support personalized search that the specifications of the user should be match by the search result. The above mentioned statements are the major complications of PageRank Algorithm. The main contribution of this paper uses a new similarity measure derived from vector space model to compute the similarity between pages based on terms, and applied it to partition the web into PageRank. It proposed a weighted PageRank algorithm, which is called Improved Candidate Transaction PageRank Algorithm (ICTPRA). It considers the relevance of a page to the given query that probably improve the accuracy of the page scoring with the short time duration. The algorithm also introduce a new PageRank, developed with the ability of filtering useless pages.

PageRank Algorithm - Random surfing model:

PageRank Algorithm - Random surfing model:

At any page, with prob. α, randomly jumping to a page

with prob. (1-α), randomly picking a link to follow

$r = A_r$, where

$$A_{ij} = \beta M_{ij} + (1-\beta)/N \tag{1}$$

$$r_i = \sum_{1 \le j \le N} A_{ij} r_j \tag{2}$$

$$r_i = \sum_{1 \le j \le N} \left[\beta M_{ij} + \left(1-\beta\right)/N \right] r_j \tag{3}$$

$$r_i = \sum_{1 \le j \le N} M_{ij} r_j + \left(1-\beta\right)/N \sum_{1 \le j \le N} r_i \tag{4}$$

$$= \sum_{1 \le j \le N} M_{ij} r_j + \left(1-\beta\right)/N \text{ ,since } \left|r\right| = 1 \tag{5}$$

$$r = \beta M r + \left[\left(1-\beta\right)/N\right]_N \tag{6}$$

4. PROPOSED WORK

The aim of the proposed work is to analyze the classifier technique abilities for an ever-increasing number of PageRank algorithms in the trained dataset. The performance is affected by the number of attributes of the feature space. This study involves PageRank, with all the possible features. Similarly, when the number of features were increased, the train dataset and test dataset was composed of its maximum size that was contained as many documents of both the classes. First of all, it is observed that architectural variables like Rank selection, algorithm, PageRank, Feature Rank selection, Performance, Feature space had often a larger impact on the performance than the choice of individual classifier. In fact, it is suitable for architectural variables, they are chosen. If the parameter settings of the classifiers get correctly optimized, the difference between the algorithms will not very large. Finally the Improved Candidate Transaction PageRank Algorithm is tested with the previous algorithms RWFIM, WEAPON, CASINO and CTRA [19 -21] each quantity in an equation.

A. Candidate Transaction Rank Algorithm (CTRA)

Candidate Transaction Rank Algorithm (CTRA) was established using by Apriori algorithm to moderate the time to get the results for the user's queries in the search engines and also in online social networks [21]. Hubs and authorities are mainly used to develop CTRA. HITS and PageRank algorithms are shared in CTRA to get better results. The CTRA is made known below. CTRA also showed the better time constraint comparing to the previous algorithm RWFIM.

B. Candidate Transaction Rank Accuracy Algorithm (CTRAA)

Now a days, Online Social Networks are the flourishing area for all categories of people in the world. It becomes very vital one for day-to-day life for real-time information sharing. It is able to search for the users to related works. Candidate Transaction Rank Accuracy Algorithm (CTRAA) was advanced by using KNN classifier algorithm to improve the accuracy and to contract the time to get superior results for the user's queries in Online Social Networks [22]. The KNN algorithm is used to recognize the benchmarks – parsing, indexing, sorting and interaction for ranking in the search results.

Recommendation of KNN algorithm based on HITS

Web Page Ranking associated to recommend the requirements are individual needs, rather than the generally popular products, so the value of the original authority and hub improvements will define extent of the hub as the associated interest. In the following CTRAA algorithm exhibited the improved accuracy and in low time constraint paralleling to WEAPON and CASINO for Online Social Networks.

C. Enhanced Candidate Transaction Rank Accuracy Algorithm (ECTRAA)

Enhanced Candidate Transaction Rank Accuracy Algorithm (ECTRAA) was developed with the help of SVM classification algorithm for the improvement of relevancy and accuracy. It is also given the better results by low time duration for the user's queries. In this ECTRAA, the Hilltop method Algorithm is used to get the better results to the users by the hyperlinks and the relevancy of the web pages using the Search Engine Optimization (SEO). Finally the ECTRAA which is shown below given better results to the users for their queries with the improved accuracy, relevancy with low time duration with the previous algorithms RWFIM, WEAPON and CASINO.

D. Candidate Transaction PageRank Algorithm (CTPRA)

AdaBoost is a very simple technique to implement for extracting the features from the large datasets in the Candidate Transaction PageRank algorithm [23]. It adjusts adaptively the errors of the weak hypotheses by Weak Learn. Measuring the differentiation between region averages at various scales, orientations and aspect ratios. However, this information is limited and needs to be boosted to perform accuracy, relevance and time classification. The integral website at location RR, TR contains the sum of the URL above and to the left of RR, TR, inclusive:

$$\left(RR,\ TR\right) = \sum_{RR \le TR, TR \le RR} i\left(RR', TR'\right) \tag{7}$$

$$A(\mathrm{RR, TR}) = \mathrm{A(RR, TR - 5) + M(RR, TR)} \tag{8}$$

$$MP(\mathrm{RR, TR}) = \mathrm{MP\ (RR\text{-}5, TR) + M(RR, TR)} \tag{9}$$

Where RR = Relevance Ratio

TR = Time Ratio
A = Accuracy
MP = Measuring Page Rank

The frame work of this algorithm is as follows

- The learner receives an example which is chosen randomly according to some fixed but unknown distribution on RRXTR.
- The learner finds a hypothesis that is consistent with most of the samples.

Training set: Given a training set of labelled examples $\{(x_1,y_1), \ldots, (x_n,y_n)\}$, estimate the prediction function f by minimizing the prediction error on the training set.

Testing set: Apply f to a new test example x and output the predicted value $y = f(x)$.

Input variables:

P: The distribution where the training examples sampling from

D: The distribution over all the training samples

Weak Learn: A weak learning algorithm to be boosted

T: The specified number of iterations

N: Total Number of webpages

Oj: Number of outgoing links from page *j*

Bi: Set of web pages pointing to web page *i*

d: damping factor (usually set as 0.85)

E. Improved Candidate Transaction PageRank Algorithm (ICTPRA)

This algorithm shows better improvement of the all the features that are selected in Candidate Transaction PageRank Algorithm. Improved Candidate Transaction PageRank Algorithm states that SVM and AdaBoost classification methods are performing better than other methods that are used. The PageRank result is 99.99% in ICTPRA.

Improved Candidate Transaction PageRank Algorithm

1. **Input: Internet** = **set of pages**
2. **For each** pagePagerank in Internet **do**
3. Pagerank authorities = 1 // Pagerankauth is the authority score of the page p
4. Pagerank hub = 1 // Pagerankhub is the hub score of the page p
5. Function Hubs and Authorities (G)
6. **For** step from 1 to k **do** // run the algorithm for k steps
7. Norm = 0
8. **For each** page Page in Internet **do** // update all authority values first
9. Pagerank authorities = 0
10. **For each** page q in Pagerank incoming Neighbors **do** // Pagerank incoming neighbors is the set of pages that link to p
11. Pagerankauthority + = q.hub
12. Norm += square (Pagerankauthorities) // calculate the sum of the squared authorities values to normalise
13. Norm = sqrt (Norm)
14. **For each** page Pagerank in Internet **do** // update the authority scores
15. Pagerankauthorities = Pagerankauthorities / Norm // normalise the authority values
16. Norm = 0
17. **For each** page rank in Internet **do** // then update all hub values
18. Pagerankhub = 0

For each page r in Pagerank outgoing Neighbors **do** // Pagerank outgoing neighbors is the set of pages that p links to
Pagerankhub += r.auth
Norm += square (Pagerankhub) // calculate the sum of the squared hub values to normalise
Norm = sqrt (Norm)
For each Pagerank in Internet **do** // then update all hub values
Pagerankhub = Pagerank hub / Norm // normalise the hub values

Nearest-neighbor and K-Nearest-Neighbor Classifier: L1 distance, $\chi 2$ distance, quadratic distance, histogram intersection PageRank website base search engine rank. KNN could deal with compiles and arbitrary decision PageRank.

Table 1
Comparison of Equations with the proposed algorithm

Algorithms	Object Web PageRank	Target WebPageRank
CTRA	$\text{highest} \sum_{se}^{web} \left[\begin{array}{l} \sum_{j}^{R} \log PR\left(p_{ser} \mid Rse; \theta r\right) \\ + \log P\left(Rse; \theta r\right) \end{array} \right]$	$\boldsymbol{\theta}_1^{T} \mathbf{x} + \boldsymbol{\theta}_0^{T} \left(1 - \mathbf{x}\right) > 0$ where $\theta_{(1,0)} = \log \frac{P\left(x_j = 1 \mid y = 1\right)}{P\left(x_j = 1 \mid y = 0\right)}$, $\theta_{(0,1)} = \log \frac{P\left(x_j = 0 \mid y = 1\right)}{P\left(x_j = 0 \mid y = 0\right)}$
CTRAA	$\text{highest} \sum_{i}^{A} \log\left(P\left(Rse \mid m, n\boldsymbol{\theta}\right)\right) + \lambda \left\lVert m\boldsymbol{\theta} \right\rVert$ website $P\left(TRr \mid mn, \boldsymbol{\theta}\right) = 1 / \left(1 + \log\left(-TRr_{se}\boldsymbol{\theta}^{T} m\right)\right)$	$C = \substack{Website \\ Rank} \quad Cp\left(Cse, R\right)$
ECTRAA	$M.Rank \quad \lambda \mathbf{p} \sum_{i}^{R} r \xi_i + \frac{1}{2} \left\lVert \boldsymbol{\theta} \right\rVert$ such that $\Pr \quad \times i^{\boldsymbol{\theta}^{T}} \mathbf{x} \geq 1 - \xi_i \forall r$	PF :T = r | r = pp | r + | PF
CTPRA	$_{PR}(x) = P(PR=1 \mid x) - P(PR=0 \mid x)$	freq(Rank) = website(SE)
ICTPRA	$\theta_{kj} = \frac{\sum_{i}^{R} \delta\left(x_{ij} = 1 \wedge y_i = k\right) + r}{\sum_{i} \delta\left(y_i = k\right) + Kr}$	$\text{Sim}(\rho) = \sum_{T}^{R} (w(v) * M(v, \rho(v))) / \sum_{T}^{A} w(v)$

w : A web page
kr : The web pages that have a link to w
P : The number of out links from wi
TR : The Time Rank
N(n) : The size of the web(m) M
PP : Number of instances covered by Search Engine
Cse : Number of instances current Search Engine
k : Number of classification PageRank
Cp : Prior PageRank
PR : PageRank
PF : Page Follow
Freq : Frequency
Pse : Page Search Engine
Rse : Rank Search Engine
r : Rank
Kr : Classification rank
x : web page
T : Time
xij : Web mining (website, webpage)

Despite its simplicity, researchers have shown that the classification accuracy of KNN could be quite strong and in many cases as accurate as those elaborated methods. KNN is slow at the classification time.

Support Vector Machines: Web mining classifiers mainly adopt this method. Margin maximization rank page and website. Kernel functions are used in histogram intersection page, generalized Gaussian Rank, pyramid match website. Multiclass pages of list are classified using neural networks, boosting, and decision trees. Table 1. gives comparison of equations for the proposed algorithm.

5. EXPERIMENTAL STUDY

The analysis of the algorithms are used to scan the database twice this paper presents an improvement on it by using the previous algorithm and the concept of partitioning. It presents a mathematical formula for selecting the PageRank. The code is implemented in Matlab 2015a and the platform is used windows 10. The architecture is used mac operating system x for calculating the running time of the algorithm. The T1014D100K dataset which was generated by the recognized IBM Quest Synthetic Data Generation is used for the evaluation of experiments [19]. The real world datasets Sina Weibo and Twitter datasets are also used to evaluate the algorithm ICTPRA [20]. The data structure which is introduced in this paper is frequent website that is used for finding out the frequent itemsets and also used for generating the conditional methods of PageRank. The analysis shows that the time consumed is decreased in the proposed algorithm of each group of transaction is less than the original web page rank. The average of the reduced time rate in the Improved Candidate Transaction PageRank Algorithm’s is 82.15%, Accuracy is 99.22% and the Relevancy is 99.77%. The memory space is reduced by using the partitioning approach where the clusters initially partitions and select one particular PageRank. It is an improvement as earlier the algorithm took exponential space but now it is reduced.

The mathematical formula for calculation the value of k is Importance score of page k = score of $k = x_k$. The experimental results of the Datasets for CTPRA and ICTPRA is explained in detail in Table. 2 and the Figure. 1 shows the difference in metaphorically. In this paper, a relation-based PageRank algorithm is used in conjunction with semantic web

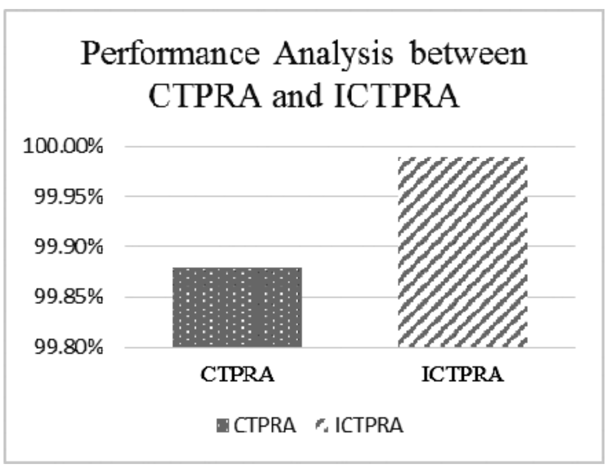

Figure. 1. Performance Analysis between CTPRA and ICTRA

search engines that simply relies on information that could be extracted from user queries and on annotated resources is proposed. The relevancy is measured as the PageRank that a retrieved resource actually contains those relations whose existence was assumed by the user at the time of query definition.

Table 2. Experimental Result for Datasets between CTPRA and ICTPRA

Datasets	*Algorithms*	*Performance*
T1014F100K	CTPRA	99.88%
SinaWeiboAnd Twitter	ICTPRA	99.99%

6. EVALUATION METHOD

The proposed work (ICTPRA) is evaluated by using precision and recall methods. It is derived from the equations for all possible techniques. Most of the techniques had the 100% of evaluation results both in precision and recall. Each and every technique that has been calculated for ICTPRA has the better precision and recall results for the search engines. The error rate is also calculated in the evaluation methods. The evaluation method of previous algorithms is given in the Table. 3 to show the improvement of the ICTPRA.

Table 3. Evaluation method

Techniques	*Equations*	*Recall*	*Precision*
Occasion	CA/PR * 100% = Rank	30%	100%
Kindliness	$\sum_{i=1}^{n} c = cn$	30%	50%
Specificity	$\sum_{i=i_0}^{n} (a_i \pm b_i) = \sum_{i=i_0}^{n} a_i \pm \sum_{i=i_0}^{n} b_i$	30%	33%
Correctness	(PR * WE * SE)/Total Rank = PR	40%	50%
Encouraging Unenthsiasic	Error 1 * Relevance/Time = Rank	60%	60%
Extrapolative Rate	Rate SE * URL Page = Rank	60%	50%
Discovery PageRank	Hub * Auth/Relevance – 1 = Rank	80%	57%
Algorithms Analysis	(Apriori + KNN + SVM + ADA) = PR	80%	50%
Error Rate	(Apriori + KNN + SVM + ADA)/ICTPRA = PageRank	60%	44%
Recall	RC = Apriori/KNN * 100%	100%	50%
Precision	PC = SVM/ADA * 100%	100%	45%
Phony Exclusion Rate	PER = RC/PC * 100%	100%	42%

CA = Correct Activist
PR = PageRank
cn = Correct Number of pages
n = Total Number of pages
ws = Web page
SE = Search Engine
RC = Recall
PC = Precision
PER = Phony Exclusion Rate

Table 4. F-Measure, Recall and Precision for Algorithms

Name of the Proposed Algorithm	*F-Measure*	*Recall*	*Precision*
CTRA	68.9%	84.3%	58.3%
CTRAA	68.0%	87.7%	55.6%
ECTRAA	68.7%	90.5%	55.4%
CTPRA	68.9%	91.2%	55.4%
ICTPRA	90.2%	93.9%	90.3%

The Table. 4 displays the F-Measure, Recall and Precision for the algorithms to get the information about the improvement in the Improved Candidate Transaction PageRank Algorithm.

7. CONCLUSION

From the above investigations, it is noted that search of the relevant information in the search engine can be a significant issue. However, this proposed work is tried to implement classifying methods like Apriori, KNN, SVM and AdaBoost are used to get better results in PageRank. Eventually the results of classification method on datasets are shown that ICTPRA has the better efficiency among other methods.

Different algorithms perform in its own way that has been depended on data collections. Some of them appears to be globally superior over the others. However, SVM and AdaBoost are good choices when all the classification algorithms has been considered in all the factors. Google is designed to be a scalable search engine. The primary goal is to provide high quality search results over a rapidly growing World Wide Web. Google employs a number of techniques to improve the search quality including PageRank, anchor text and proximity information. Furthermore, Google is a complete architecture for gathering web pages. Alike, indexing them and performing search queries over them.

A large-scale web search engine is a complex system and remains to do PageRank in an effective manner is little challenging. The immediate goals are to improve search efficiency and to scale the 100 million web pages approximately. Some simple improvements are needed for indexing, include query caching, smart disk allocation and sub-indices. Smart algorithms are to be decided, what old web pages should be re-crawled and what new ones should be crawled. The proposed work has fulfilled this goal. One promising area of research is using proxy caches to build search databases, since they are demand driven. The proposed algorithm has been added simple features supported by commercial search engines like Boolean operators negation and stemming. However, other features are just starting to be explored such as relevance, feedback. The ICTPRA supports the user context and the result summarization. This is also extended with the use of link structure and link website. Simple experiments indicate PageRank could be personalized by increasing the weight of a user's home page or bookmarks. As for link text, the experiments are done by using website links in addition to the link test itself. A web search engine is a very rich background for research ideas. In future, the ICTPRA has been improved with more features that the researcher need.

REFERENCES

[1] Themistoklis Marvridis and Andreas L. Symeonidis, "Identifying valid search engine ranking factors in a web 2.0 and web 3.0 context for uilding efficient SEO mechanisms," *Elsevier, Eningeering Applications of Artificial Intelligence*, **41**, 75-91, 2015.

[2] Changsung Kang, Dawei Yin, Ruiqiang Zhang, Nicolas Torzec, Jianzhan He and Yi Chang, "Learning to rank related entities in Web search," *Elsevier, Neurocomputing,* **166**, 309-318, 2015.

[3] Yan Chen, Xianoming Zhang li and Jun-Ping Ng, "Search enigne reinforced semi-supervised classification and graph-based summarization of microblogs," *Elsevier, Neruocomputing,* 152, 274-286, 2015.

[4] Ali Seyfi, Ahmed Patel and Joqauim Celestino Junior, "Emprical evaluation of the link and content-based focused Treasure-crawler," *Elsevier, Computer Statndards & Interfaces,* **44**, 54-62, 2016.

[5] Hebert Perez-Roses, Francesc Sebe and Josep Maria Ribo, "Endorsement deduction and ranking in Social Networks," *Elsevier, Computer Communications,* **73, Part B**, 200-2010, 2016.

[6] Di Jiang, Kenneth Wai-Ting Leung, Lingxiao Yang and Wilfred Ng, "TEII: Topic Enhanced Inverted Indes for top-k document retrieval," *Elsevier, Knowledge-Based Systems,* **89**, 349-358, 2015.

[7] Jihyun Lee, Jun-Ki Min, Alice Oh and Chin-Wan Chung, "Effective ranking and search techniques for Web resources considering semantic relationships," *Elsevier, Information Processing and Management,* **50**, 132- 155, 2014.

[8] Sergey Brin and Lawrence Page, "Reprint of: The anatomy of a large-scale hypertextual web search engine," *Elsevier, Computer Networks,* **56**, 3825-3833, 2012.

[9] Jon M.Kleinberg, "Authoritative Sources in a Hyperlinked Environment," *ACM*, 1-34, 1998.

[10] Dongsheng Duan, Yuhua Li, Ruixuan Li, Rui Zhang, Xiwu Gu and Kunmei Wen, "LIMTopic: A Framework of Incorporating Link Based Importance into Topic Modeling," *IEEE Transactions on Knowledge and Data Engineering,* **26(10)**, 2493-2506, 2014.

[11] Zhe Yao, Philip Mark and Michael Rabbat, "Anomaly Detection Using Proximity Graph and PageRank Algorithm," *IEEE Transaction on Information Forensics and Security*, **7(4)**, 1288-1300, 2012.

[12] MarcClasesen, FrankDeSmet, Johan A.K.Suykens and BartDeMoor, "A robustensemble approach to leanarn from positive and unlabeled data using SVM basemodels," *Elsevier, Neurocomputing,* **160**, 73-84, 2015.

[13] Kumar Ravi and Vadlamani Ravi, "A Survey on opinion mining and sentiment analysis: Tasks, approaches and applications," *Elsevier, Knowledge-Based Systems,* 1-72, 2015.

[14] Wilco W.M.Fleuren and Wynand Alkema, "Application of text mining in the biomedical domain methods," *Elsevier, Methods,* **74**, 97-106, 2015.

[15] Noraini Seman and Nursuriati Jamil, "Blending Sentence Optimization Weights of Unsupervised Approaches for Extrctive Speech Suumarization," *Elsevier, Procedia Computer Science,* **51**, 620-629, 2015.

[16] Shufeng Xiong and Donhong Ji, "Query-focused multi-document summarization using hypergraph-base ranking," *Elsevier, Infomration Processing and Management,* 1-12, 2016.

[17] Ben Brahim A, Liman M, "Robust ensemble feature selection for high dimensional datasets," ***International Conference on High Performance Computing and Simulation, IEEE***, 151-157, 2013.

[18] Tamir Tassa and Dror J.Cohen, "Anonymization of Centralized and Distribtued Social Networks by Sequential Clustering," *IEEE Transactions on Knowledge and data engineering,* **25(2)**, 311-324, 2013.

[19] Jerry Chun-Wei Lin, Wensheng Gan, Philippe Fournier-Viger and Tzung-Pei Hong, "RWFIM: Recent weighted-frequent itemsets mining," *Elsevier, Engineering Applications of Artificial Intelligence,* **45**,18-32, 2015.

[20] Li Hui, SHEN Bingqing, CUI Jaingtao and MA Jianfeng, "UGC-Driven Social Influence Study in Online Micro-Blogging Sites," *ICT Management, China Communications,* 141-151, 2014.

[21] Srinaganya.G and Dr.J.G.R.Sathiaseelan, "Discovering a Solution for Web Hubs using Candidate Transaction Rank Algorithm," *International Journal of Applied Engineering Research,* **10(82)**, 137-141, 2015.

[22] Srinaganya.G. and J.G.R.Sathiaseelan, "Candidate Transaction Rank Accuracy Algorithm for Online Social Networks," *Middle-East Journal of Scientific Research,* **23(12)**, 2868-2874, 2015.

[23] Srinaganya.G. and J.G.R.Sathiaseelan, "Candidate Transaction PageRank Algorithm using AdaBoost for Web Search Engines," *World Applied Sciences Journal,* **34(1)**, 128-139, 2016.