Journal of Theoretical and Applied Information Technology
30th June 2017. Vol.95. No 12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

BIO-MEDICAL LITERATURE RETRIEVAL FROM MIXED LITERATURE BANK WITH THE AID OF MULTI KERNEL FUZZY C-MEANS TECHNIQUE (MK-FCM)

[1]THAYYABA KHATOON MOHAMMED, [2]DR. A.GOVARDHAN,[3]DR.H.S.SAINI, [4]DR RISHI SAYAL

[1]Research scholar, JNTUH
[2]EC member,Professor&Principal,dept of CSE,JNTUHCEH
[3]Professor,Dept Of CSE,GNITC.
[4]Professor,Dept Of CSE,GNITC
Email: thayyabakhatoonphd@gmail.com

ABSTRACT

In the apparent, biomedical literature will be retrieved starting with the mixed literature articles, utilizing Multi-Kernel Fuzzy c Means (MK-FCM) clustering technique. Literature retrieval or document retrieval is an action that utilization professional techniques for medically examine papers retrieval, the report card also different information to move forward research and practice. In the work, the biomedical query article is made also preprocessing system may be utilized and there would two sorts of preprocessing to be specific stop words removal and stemming. Then afterward those executions for preprocessing the retrieval techniques, for example, vector space modeling and retrieval modeling is utilized. In the vector space modeling the term frequency and inverse frequency measure may be broken down after that on retrieval modeling the SMART and BM25 standard models used to furthermore retrieve the literature. Behind this Multi-Kernel Fuzzy c Means (MK-FCM) technique is utilized for clustering the literature. In this differentiate kernel function named as FCM, Linear FCM, Quadratic FCM and Composite kernel function will be investigated to several measuring tests. Those biomedical query article databases where brain tumor, breast cancer, kidney stone and neovascularization is clustered. For single, randomly two, randomly three and overall four types of input biomedical documents the precision of the composite kernel function is 90% and recall of composite kernel function is 87% were compared with different strategies and it progresses the level best. By using these techniques, the retrieval performed better from other techniques.

Keywords: *Biomedical literature, Multi-Kernel Fuzzy c Means (MK-FCM) clustering, Composite kernel function, Linear FCM and Quadratic FCM.*

1. INTRODUCTION

Literature retrieval (LR) is the undertaking of discovering records of an unstructured nature that fulfill a data require from inside substantial accumulations. LR can cover different and heterogeneous sorts of information and data issues, for example, web looking and crawling, personal LR, area particular search.[1] With expanding digitization of libraries, literature works of different neighborhood dialects from various states, are being reestablished in digitized shape by applying inventive transliterations and encoding formats.[2] In numerous cases, the full content is not accessible to the analyst by any means; the literature databases contain just bibliographic data and edited compositions. Last ones experience the ill effects of impediments of data compression and convolution forced by a word limit.[3] Information retrieval works by ordering content and after that chooses helpful data. Data Extraction concentrates on removing important truths while information retrieval chooses pertinent document.[4] The related information retrieval instruments are the primary hotspot for clients to get the data rapidly. Due to the 'information sources explosion', it is troublesome for clients to get the data as exact as possible.[5] Information retrieval is the errand of recovering important data from the PDF report. The framework retrieves distinctive sort of information and data from entire PDF files.[6] The most widely recognized approach here

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

comprises in giving clients after they select a specific output with an arrangement of related archives, retrieved on the premise of some similitude measure between those documents.[7] Nowadays, experts require more propelled record indexation designs and advanced natural language processing (NLP) methods to handle the present data development on the Web.[8] There are numerous open issues because of troubles of managing non-structured texts portrayed in natural language with the specificities of medical information.[9] PubMed supports keyword and constraint queries. Be that as it may, a keyword query regularly gives back a not insignificant rundown of hits. For instance, the keyword "Parkinson's disease" retrieves more than seventy thousand articles.[10] Every report that matches the posted keywords in any of the asked for hunt fields (e.g. title, keywords or abstract) is viewed as a hopeful. In any case, it is not minor for the client to represent query in a manner that the keywords (single words or sets of words anticipated that would co-happen together) don't bring consideration over reports that are not associated with the subject of their interest.[11] Some elements that are frequently considered for sentence choice are word frequency, title words, cue words, sentence area and sentence length. These elements regularly increment the office of a sentence for incorporation in summary.[12] Clustering is a method for classifying patterns or questions in such a way, to the point that specimens of the same cluster are more practically identical to each other than tests having a place with different clusters. There are two fundamental bunching approaches: the hard clustering procedure and the fuzzy clustering technique.[13] Clustering is broadly utilized as a part of different problem-solving and decision-making applications, for example, document retrieval, marketing research, genotype assignment, insurance fraud identification, image segmentation, data analysis and city-planning.[14] The aim of this paper is retrieving the literatures from the database based on given bio medical query keyword. The proposed classifier creates a nonlinear acumen work in the yield space and has the better execution of report recognition.[15] For instance, in document classification, fuzzy co-clustering allows any record and word to have a place with more than a single co-cluster. Fuzzy co- clustering is appropriate for clustering complex information sorts as multi-dimensional, multi-highlights, and of expansive size.[16]

2. LITERATURE REVIEW

Rafal Lancucki *et al.* [17] 2016, had proposed Standard ways to deal with learning extraction from biomedical literature concentrate on information retrieval from edited compositions freely accessible in medical databases like PubMed. Be that as it may, for some exploration themes the pre-choice of the sufficiently little arrangement of the archives can be exceptionally troublesome or even impossible. Another issue comes from huge changeability of the retrieved arrangements of productions while changing keywords in web indexes. In this paper address both of these issues by proposing an algorithm and an execution fit for taking a shot at the full content articles. Display an information retrieval framework with a choice of discrete segments of full messages of papers and manage based the web search tool. Show that in some examination our answer can give much preferable comes about over discovering reports just by keywords and edited compositions.

Stephen Fitchett *et al.* [18] 2015, had recommended Retrieving records on PCs is a central segment of the connection, yet there is shockingly minimal experimental information describing how it was completed in practical settings. Created programming, called File Monitor, to progressively record clients' document recovery exercises, including information portraying the files retrieved and the apparatuses used to recover them. We then sent the framework in a four-week log investigation of 26 members ‘actual file retrievals on their PCs. Follow-up meetings contextualized the discoveries. Results are exhibited in two segments concentrating on the records (the number of files, patterns of revalidation, file types and so forth.) and on the interface systems used to retrieve them (file browsers, search tools, ‘recent files’ lists and so on.). We finish up by talking about suggestions for the plan of cutting edge document recovery interfaces.

Poonam Yadav [19] 2014, Literature presents diverse calculations for web archive bunching helpful for data recovery. In this work, Document-Document similitude grid and Multiple-Kernel Fuzzy C-Means Algorithm-based web document clustering are produced for information retrieval. At to start with, web records are perused and introductory pre-preparing is connected to remove the vital words. At that point, include space is built utilizing watchwords and its recurrence. Accordingly, report to archive similitude lattice is built utilizing the comparability measure, called semantic retrieval measure (SR). The measure considers four unique criteria, for example, the probability of occurrence in the document, probability of an event in the main report, the probability of an event in the second document and probability of an event in both equivalent words set. Taking into account this measure, D-D framework is figured to do the last gathering utilizing Multiple-Kernel Fuzzy C-Means Algorithm. The experimentation is finished with 100 web records and the outcomes are assessed with accuracy and entropy.

Paavo Nieminen *et al.* [20] 2013, had recommended the learning disclosure procedure to the mapping of current themes in a specific field of science. Inspired by how articles frame clusters and what are the substance of the discovered clusters. A system including web scratching, keyword extraction, dimensionality diminishment and clustering utilizing the dispersion map algorithm is introduced. Utilize freely accessible data about articles in high-impact journals. The strategy ought to be useful to professionals or researchers who need to review late research in a field of science. As a contextual analysis, we outline themes in information mining literature in the year 2011.

Hsin-Chien Huang *et al.* [21] 2012, had foreseen fuzzy c-means was a prominent soft clustering technique; its viability was largely restricted to spherical clusters. By applying kernel tricks, the kernel fuzzy c-means algorithm endeavors to address this issue by mapping information with nonlinear connections to suitable element spaces. Kernel combination or selection is crucial for viable kernel clustering. Shockingly, for most applications, it is difficult to locate the right mix. Propose a multiple kernel fuzzy c-means (MKFC) algorithm that augments the fuzzy c-means algorithm with a multiple kernel learning setting. By consolidating multiple kernels and naturally conform the kernel weights, MKFC is more insusceptible to ineffectual portions and unessential elements. This settles on the selection of kernels less pivotal. What's more, we demonstrate various multiple kernel k-means (MKKM) to be an exceptional instance of MKFC.

3. PROPOSED METHOD

The essentialness goal of this proposed work is retrieving the literature from the database in light of given biomedical query article. The biomedical query article is taken for preprocessing in these two distinct sorts to be specific stop words removal and stemming are performed in preprocessing. After this procedure, the retrieval techniques are utilized for retrieving the archives, for example, vector space modeling and retrieval modeling. In vector space modeling, the term frequency and inverse frequency measure is broke down then in retrieval modeling the SMART and BM25 standard models utilized and retrieve the literature. Here, consolidate the grouping strategy as multi-kernel based fuzzy c means (MK-FCM) technique. Essentially, in the work applying separate kernel function (linear and quadratic) in FCM, a later hybrid of aforementioned is incorporated to form composite kernel FCM. This proposed strategy works bosses in this stage for high characterization rate. Taking after, measures are used to test its performance several metrics. This complete work will actualize in the working stage MATLAB. Figure 1 shows the overall flowchart for proposed method.

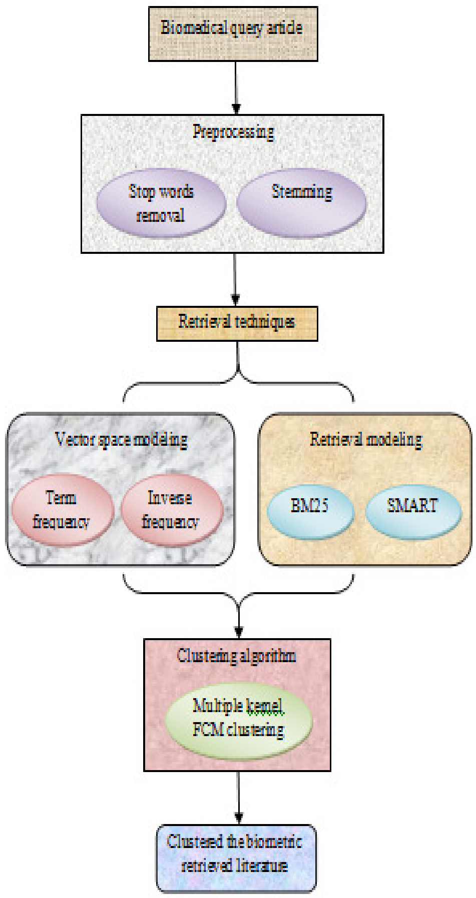

Figure.1 Overall Flowchart For Proposed Method

3.1 Biomedical Query article

In the work, using biomedical query article database named as a Brain tumor, Breast cancer, Kidney stone, and Neovascularization. These types of biomedical articles are used for preprocessing process below.

3.2 Preprocessing

Raw data is amazingly powerless against clamor, missing qualities, and irregularity. The standard of information influences the data mining comes about. To enhance the standard of the data and subsequently of the mining results is pre-processed therefore on enhancing the strength and straightforward the mining technique. Data preprocessing is one among the preeminent basic strides in a data mining process that arrangements with the readiness and change of the underlying dataset. The two techniques utilized for preprocessing the given reports are:

- Stop words Removal
- Stemming

3.2.1 Stop words Removal

The stop words removal approach is utilized to wipe out the undesirable words, for example, sometime recently is, an, a, the, become, then, they, there, that, them, and so forth.

3.2.2 Stemming

The stemming algorithm is utilized to dispense with the stemming words and to recognize the root words. The stemming words which are closure with ed, ion, ing. Stemming has not connected for the reasons of i) losing the context of search, ii) may reduce precision and iii) cannot be applied to proper nouns. To have an unmistakable perspective of pertinence, rather than utilizing given arrangement of query and significance judgments, we have made our arrangement of queries from 'titles' of 4 documents, forming a query set of 4 entities.

3.3 Retrieval technique

In this retrieval technique, there are two models namely

- Vector space modeling
- Retrieval modeling

3.3.1 Vector space modeling

For the vector model, the weight $W_{u,v}$ connected with a couple (h_u, c_v) is positive and non-binary. Facilitate, the file terms in the query are likewise weighted. Let W_u, k be the weight connected with the match (h_u, k) where $W_u, k \geq 0$. At that point, the query vector $\vec{k}$ is characterized as $\vec{k} = (W_{1k}, W_{2k}, ...W_{tk})$ where t is the aggregate number of file terms in the framework. The vector for a document $\vec{c}_v$ is represented by $\vec{c}_v = (W_{1v}, W_{2v}, ...W_{tv})$.

The vector show proposes to assess the degree of similarity of the document $\vec{c}_v$ with respect to the query k as the connection between's the vectors

$\vec{c}_v$ and $\vec{k}$. This relationship can be measured by the cosine of the edge between these two vectors as,

$$sim\,(c_v\,,k) = \frac{\vec{c}_v\,.\vec{k}}{|\,\vec{c}_v\,|\times|\,\vec{k}\,|} \quad (1)$$

$$= \frac{\sum_{u=1}^{t} W_{u,v}\times W_{u,k}}{\sqrt{\sum_{u=1}^{t} W_{u,v}^2}\times\sqrt{\sum_{v=1}^{t} W_{u,k}^2}} \quad (2)$$

Where $|\,\vec{c}_v|$ and $|\,\vec{k}\,|$ are the standards of the document and query vectors. The component $|\,\vec{k}\,|$ does not influence the positioning (i.e., the ordering of the documents) since it is the same for all documents. The factor $|\,\vec{c}_v|$ gives standardization the space of the documents.

Since $W_{u,v}\ \geq 0$ and $W_{u,k}\ \geq 0$, $sim\,(k, c_v)$ differs from 0 to +1, the vector demonstrates positions the documents as indicated by their level of comparability to the query. A document may be retrieved regardless of the possibility that it coordinates the query just somewhat.

In the vector space model, the frequency of a term d_u inside a document c_v alluded to as the tf component and gives one measure of how well that term portrays the document contents. Moreover, the inverse of the frequency of a term d_u among the documents in the accumulation alluded to as the inverse document frequency or the ucf element.

The normalized frequency $f_{u,v}$ of term d_u in document c_v is given by

$$f_{u,v} = \frac{freq_{u,v}}{\max_k freq_{k,v}} \quad (3)$$

Where the maximum is computed over all terms, which are mentioned in the text of the document c_v. The ucf_u inverse document frequency for d_u , be given by

$$ucf_u = \log\frac{N}{n_u} \quad (4)$$

The best known term-weighting schemes use weights which are given by

$$W_{u,v} = f_{u,v}\times\log\frac{N}{n_u} \quad (5)$$

Such term-weighting strategies are called $tf-ucf$ schemes

The accomplishments of vector space demonstrate lie in its term-weighting plan, its partial matching strategy, and similarity measure. Common autonomy of list terms has said to be the inconvenience of vector space model yet essentially; the thought of term conditions is not productive. From the examination outcome in the field, it appears that the vector model is either predominant or nearly in the same class as the known choices.

3.3.2 Retrieval modeling

Two standard IR models with term smoothing are utilized for the retrieval bias analysis. They are the OKAPI retrieval model BM25 and SMART.

3.3.2.1 BM25

Okapi BM25 is ostensibly a standout amongst the most imperative and broadly utilized information retrieval model. It is a probabilistic capacity and nonlinear blend of three key characteristics of a document: term frequency $t_{t,c}$ document frequency cf_t and the document length $|\,c\,|$. The viability of BM25 is controlled by two parameters d and b. These parameters control the commitments of term frequency and document length.

$$BM25(c,k)=\sum_{t=k}\log\frac{|D|-cf_t+0.5}{cf_t+0.5}\frac{tf_{t,c}(d+1)}{tf_{t,c}+c(1-b+b\frac{|c|}{|\bar{c}|})} \quad (6)$$

$|c|$ is the average document length in the collection from which the documents are drawn. d and b are two parameters, and they are used with d=2.0 and b=0.75.

3.3.2.2 SMART

The System for Manipulating and Retrieving Text (SMART) is a retrieval model in information retrieval. It is based on the Vector Space Model.

$$SMART(c,k)=\sum_{t\in k}(W_c*W_k) \quad (7)$$

$$W_c=\frac{1+\log(tf_{t,c})}{1+\log(ajtf)}*\frac{1}{0.8+0.2\frac{itf}{pujot}} \quad (8)$$

$$W_k=(1+\log(tf_{t,c}))*\log\frac{|D|+1}{cf_t} \quad (9)$$

$ajtf$ represents the average number of occurrences of each term in the c, itf is the number of unique terms in c, and pivot represents the average number of unique terms per document.

3.4 Multiple Kernel Fuzzy C-Means Clustering Algorithm

Define a nonlinear map as Φ: u → Φ(u) ∈ F, where u∈ U. U denotes the data space, and F the transformed feature space with higher or even infinite dimension. KFCM minimizes the following objective function

$$k(u,v)=\frac{[u^s\ v+c]+[1-\frac{\|u-v\|^2}{\|u-v\|^2+c}]}{2} \quad (10)$$

Where this is also known as composite function.

Where,

$k(u,v)=u^s\ v+c$ is a linear kernel function and $k(u,v)=1-\frac{\|u-v\|^2}{\|u-v\|^2+c}$ is a rational quadratic function.

3.4.1 Linear kernel function

The linear kernel will be those simplest kernel functions. It may be provided for by those inward items $<u,v>$ also an optional constant C. Kernel algorithms utilizing a linear kernel are frequently equal to their non-kernel counterparts,

$$k(u,v)=u^S\ v+c \quad (11)$$

3.4.2 Rational Quadratic kernel function

The rational quadratic kernel will be less computationally escalated consideration over those Gaussian kernels also could make utilized as an elective when utilizing that Gaussian turns into as well exorbitant.

$$k(u,v)=1-\frac{\|u-v\|^2}{\|u-v\|^2+c} \quad (12)$$

It may be worth to note that in spite of the fact that Eqn. (10) would determine to utilize those composite function, utilize different functions fulfilling $K(u,v)=1$ in Eqs. (11)~ (12) to genuine provisions for example, such that the over indicated linear kernel functions and rational quadratic functions.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

That $k(u,v)$ takes as those Composite, Linear and Quadratic function likewise those kernel functions individually. Since to inadequate dataset, an information point with missing parts is probable on transform under an outlier, that algorithm in light of MKFCM to cluster inadequate information will be for the great possibility.

3.4.3 Fuzzy c-means algorithm

In the FCM algorithm, $U=\{U\,i\}=1$ is a set of N feature vectors. The fuzzy clustering algorithm maps data U into C fuzzy clusters.

$$J(x,y)=\sum_{i=1}^{N}\sum_{c=1}^{C} x_{ic}{}^{m}\, d^{2}(U_i, Y_c) \tag{13}$$

$$\sum_{c=1}^{c} x_{ic}=1\forall i, x_{ic}\geq 0\forall i, c and \sum_{i=1}^{N} > 0\forall c$$

Matrix X represents a fuzzy partition and x_{ic} denotes the degree that U belongs to the cluster C, as shown in Equation (13). In Equation (13), m is the fuzzy degree. Thus, when m = 1, the fuzzy C-Means algorithm is equal to the normal k-means algorithm. y_c represents, the C cluster centers.

$$J_\lambda(x,y)=\sum_{i=1}^{N}\sum_{c=1}^{C} x_{ic}{}^{m} d^2(U_i,Y_c)+\lambda\left(\sum_{c=1}^{C} x_{ic}-1\right) \tag{14}$$

The algorithm utilizes the Lagrange multipliers (14) to optimize the parameters.

$$X_{ic}= 1\Big/ \sum_{c'=1}^{C}\left(\frac{d(u_i,y_c)}{d(u_i,y_{c'})}\right)^{\frac{2}{m-1}} \tag{15}$$

$$Y_c=\sum_{i=1}^{N} x_{ic}{}^{m} U_i \Big/ \sum_{i=1}^{N} x_{ic}{}^{m} \tag{16}$$

The algorithm updates x_{ic} and y_c using (15) and (16). This process is repeated until $\left\| X^{(t)} - X^{(t-1)} \right\| < \varepsilon$, that ε is the termination criterion.

4. RESULT AND DISCUSSION

In this segment, those biomedical query articles will be utilized for retrieving the articles to be a specific Brain tumor, Breast cancer, Kidney stone, and Neovascularization. That MKFCM clustering algorithm will be utilized for clustering the literature in this differentiate kernel functions are handled FCM, Linear kernel, Quadratic kernel, and Composite functions. These functions are performed and measure the test as True Positive (TP), True negative (TN), False Positive (FP), False Negative (FN), True Negative Rate (TNR), True Positive (TP), False Negative Rate (FNR), False Discovery Rate (FDR), F1 Score, Precision and Recall.

4.1 Biomedical query articles for precision

In figure 2 the biomedical query articles for precision using different clustering techniques has been shown below. In this precision for a given query article the ratio of an accurately retrieved article from the number of retrieved related article. This sort of measure reveal that for a given brain tumor article FCM reveals 65%, Linear FCM reveals 73%, Quadratic FCM reveals 74% and Composite function reveals 89%. For a given breast cancer query article FCM reveal 65%, Linear FCM reveals 75%, Quadratic FCM reveals 75% and Composite function reveals 90%.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

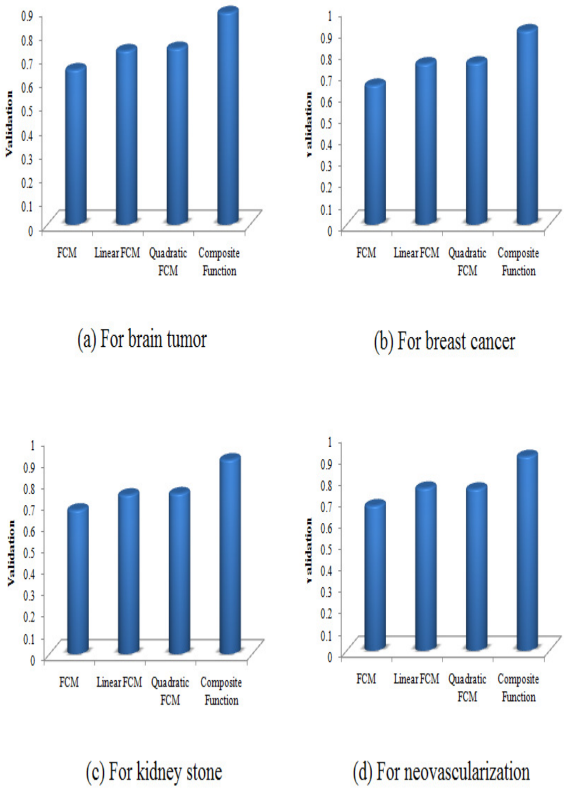

Figure.2 Biomedical Query Articles For Precision Using Different Clustering Techniques

For a given kidney stone query article FCM reveal 67%, Linear FCM reveals 74%, Quadratic FCM reveals 74% and Composite function reveal 90%. For a given neovascularization query article FCM reveal 67%, Linear FCM reveal 75%, Quadratic FCM reveal 75% and Composite function reveal 90%. From this, the composite function technique is performed better when comparing with other techniques.

4.2 Biomedical query articles for recall test

In figure 3 the biomedical query articles for recall, using different clustering techniques has been shown below. In this recall for a given query article, the ratio of the correctly retrieved article to the total number of the article related to the query article in the database. This sort of measure reveal that for a given brain tumor query article FCM reveal 61%, Linear FCM reveal 73%, Quadratic FCM reveal 75% and composite kernel function reveals 91%. For a given breast cancer query article FCM reveal 63%, Linear FCM reveal 74%, Quadratic FCM reveal 74% and composite kernel function reveal 92%.

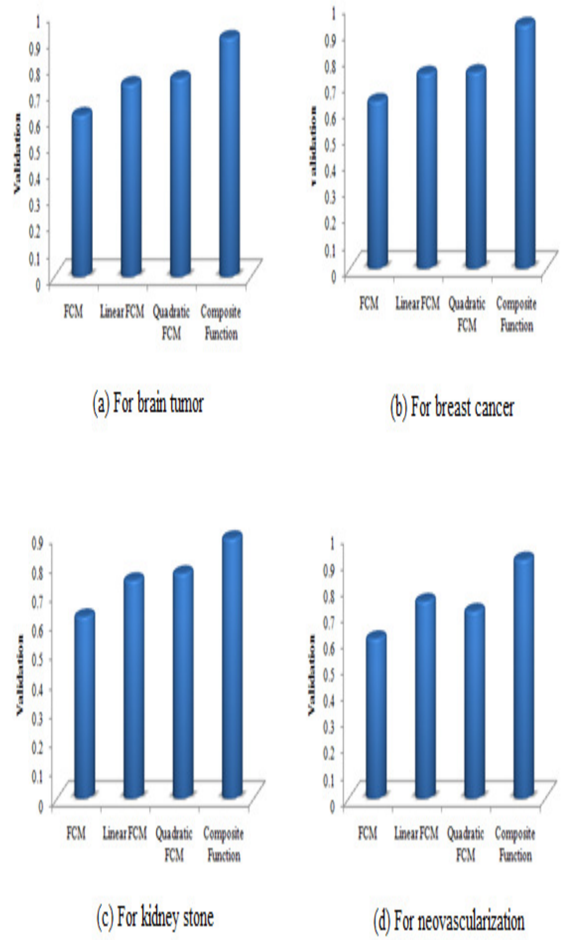

Figure.3 Biomedical Query Articles For Recall Using Different Clustering Techniques

For a given kidney stone query article FCM reveal 62%, Linear FCM reveal 74%, Quadratic FCM reveal 77% furthermore composite function reveal 89%. For a given neovascularization query article FCM reveal 61%, Linear FCM reveal 75%, Quadratic FCM reveal 71% and composite function reveal 91%. Starting with this the composite function technique will be performed better when comparing with other techniques.

Journal of Theoretical and Applied Information Technology
30th June 2017. Vol.95. No 12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table.1 The Overall Average Of Biomedical Query Articles For Precision And Recall

Sl. No	Measuring test	FCM	Linear FCM	Quadratic FCM	Composite function
1.	Precision	66%	74%	75%	90%
2.	Recall	62%	74%	74%	91%

From table 1 the overall average about biomedical query articles for precision and recall measuring test need to be demonstrated over. For a randomly selected input query document the average for precision FCM expose 66%, Linear FCM expose 74%, Quadratic FCM expose 75% and Composite function expose 90%. The average for recall the FCM expose 62%, Linear FCM expose 74%, Quadratic FCM expose 74% also composite function expose 91%. That composite function is improved when compared with other techniques.

4.3 Performance Measures For Dual Biomedical Query Article

Below figure shows, the performance measures along with different techniques while applying two biomedical query articles. The two input biomedical documents are utilized in discovering several measuring metrics. Randomly chosen two query articles are given to the system for analyzing the performance measure, here 10 validations taken place to reveal the mean value attain from different algorithms. True Positive (TP) FCM reveals 8.75, Linear FCM reveals 9.5, Quadratic FCM reveal 9.5 and Composite function reveal 13.25. The average for True Negative (TN) FCM reveal 11.75, Linear FCM reveals 12.125, Quadratic FCM reveal 12.875 and Composite function reveal 14.625. The average for False Positive (FP) FCM reveal 2.75, Linear FCM reveals 2.375, Quadratic FCM reveal 2.625 and Composite function reveal 0.625. The average for False Negative (FN) FCM reveal 5.75, Linear FCM reveals 5, Quadratic FCM reveal 6 and Composite function reveal 2.

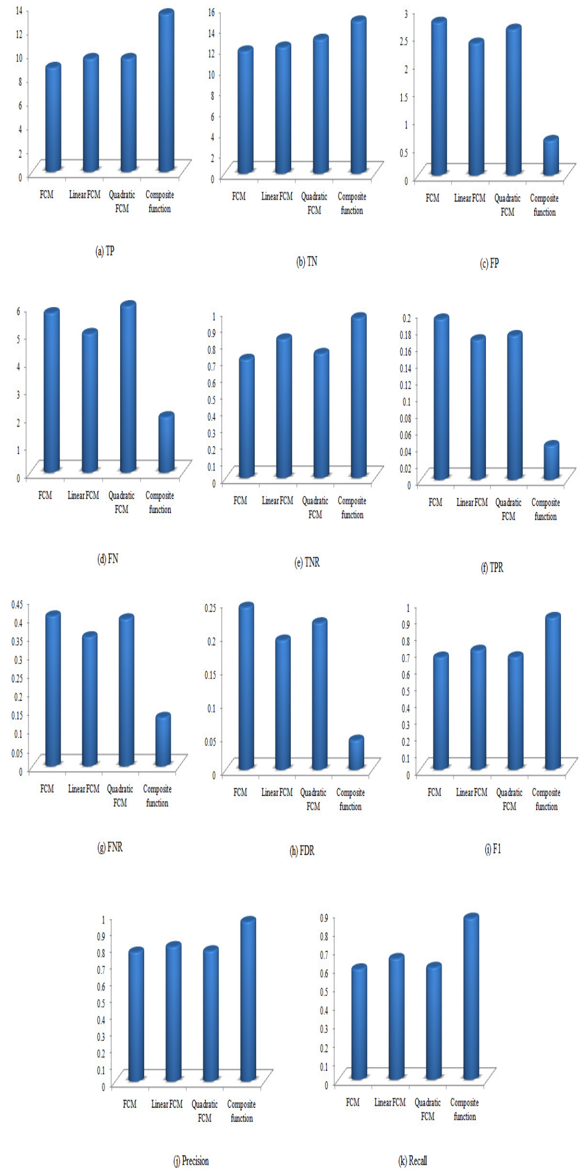

Figure.4 Average Value Attains From Performance Measure For Different Techniques By Applying Two Biomedical Query Articles

The average for True Negative Rate (TNR) FCM reveal 71%, Linear FCM reveals 83%, Quadratic FCM reveal 74% and Composite function reveal 95%. The average for True Positive Rate (TPR) FCM reveals 19%, Linear FCM reveals 16%, Quadratic FCM reveal 17% and Composite function reveals 4.09%. The average for False Negative Rate (FNR) FCM reveals 40%, Linear FCM reveals 34%, Quadratic FCM reveal 39% and Composite function reveal 13%. The average for False Discovery Rate (FDR) FCM reveal

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

24%, Linear FCM reveals 19%, Quadratic FCM reveal 21% and Composite function reveal 4.41%. The average for F1 Score FCM reveal 67%, Linear FCM reveals 71%, Quadratic FCM reveal 67% and Composite function reveal 90%. The average for Precision FCM reveal 77%, Linear FCM reveals 80%, Quadratic FCM reveal 78% and Composite function reveal 95%. Finally, the average for Recall FCM reveal 59%, Linear FCM reveal 65%, Quadratic FCM reveal 60% and Composite function reveal 86%.

4.4 Performance Measures For Three Biomedical Query Articles

Figure 5 shows, the performance measures along with different techniques while applying three biomedical query articles. The three input biomedical documents are utilized in discovering several measuring metrics. Randomly chosen three query articles are given to the system for analyzing the performance measure, here 10 validations taken place to reveal the mean value attain from different algorithms. True Positive (TP) FCM exposes 8.99, Linear FCM exposes 9.77, Quadratic FCM expose 9.77 and Composite function expose 13.11. The averages for True Negative (TN) FCM expose 24.5, Linear FCM exposes 24.7, Quadratic FCM expose 26.2 and Composite function expose 28.6. The averages for False Positive (FP) FCM expose 4.33, Linear FCM exposes 4.11, Quadratic FCM expose 4.66 and Composite function expose 1.33. The averages for False Negative (FN) FCM expose 5.44, Linear FCM exposes 4.66, Quadratic FCM expose 5.66 and Composite function expose 1.88.

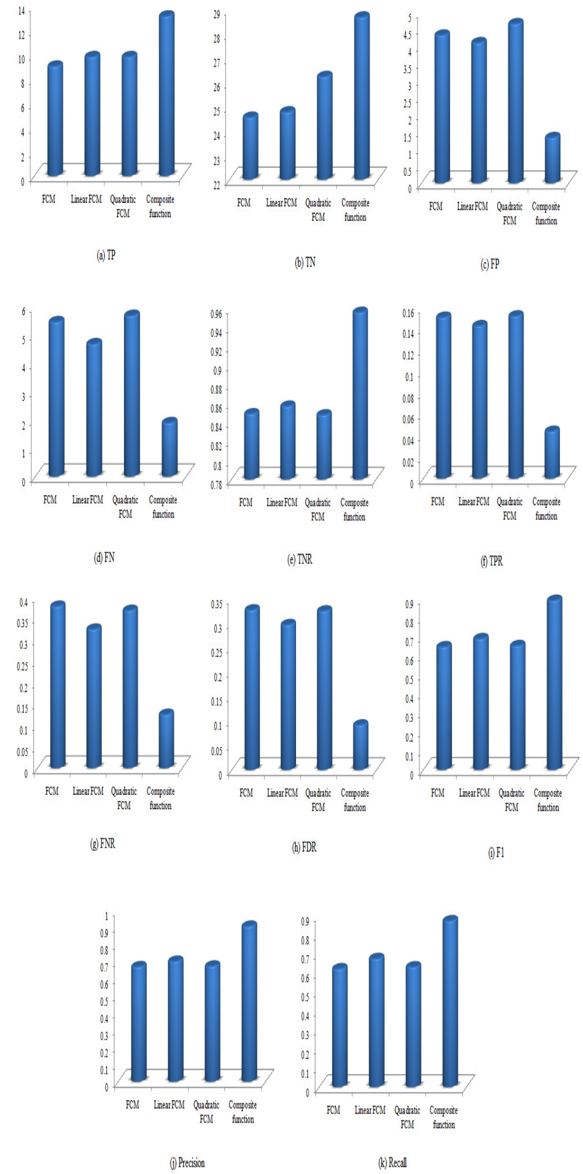

Figure.5 Average Value Attains From Performance Measure For Different Techniques By Applying Three Biomedical Query Articles

The averages for True Negative Rate (TNR) FCM expose 84%, Linear FCM exposes 85%, Quadratic FCM expose 84% and Composite function expose 95%. The averages for True Positive Rate (TPR) FCM expose 15%, Linear FCM exposes 14%, Quadratic FCM expose 15% and Composite function expose 4.4%. The averages for False Negative Rate (FNR) FCM expose 37%, Linear FCM exposes 32%, Quadratic FCM expose 36% and Composite

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

function expose 12%. The averages for False Discovery Rate (FDR) FCM expose 32%, Linear FCM exposes 29%, Quadratic FCM expose 32% and Composite function expose 9.1%. The averages for F1 Score FCM expose 64%, Linear FCM exposes 68%, Quadratic FCM expose 65% and Composite function expose 89%. The averages for Precision FCM expose 67%, Linear FCM exposes 70%, Quadratic FCM expose 67% and Composite function expose 90%. Finally, the averages for Recall FCM expose 62%, Linear FCM expose 67%, Quadratic FCM expose 63% and Composite function expose 87%.

4.3 Performance Measures For Four Biomedical Query Articles

Below figure 6 shows, the performance measures along with different techniques while applying four biomedical query articles. The four input biomedical documents are utilized in discovering several measuring metrics.Randomly chosen four query articles are given to the system for analyzing the performance measure, here 10 validations taken place to reveal the mean value attain from different algorithms. True Positive (TP) for FCM is 8.75, Linear FCM is 9.5, Quadratic FCM is 9.5 and Composite function is 13.25 then the True Negative (TN) for FCM reveal 37.75, Linear FCM reveals 38.5, Quadratic FCM reveal 40.5 and Composite function reveal 43.75. The average for False Positive (FP) FCM reveal 5.75, Linear FCM reveals 5, Quadratic FCM reveal 6 and Composite functions reveal 2. The average for False Negative (FN) FCM reveal 5.75, Linear FCM reveals 5, Quadratic FCM reveal 6 and Composite functions reveal 2.

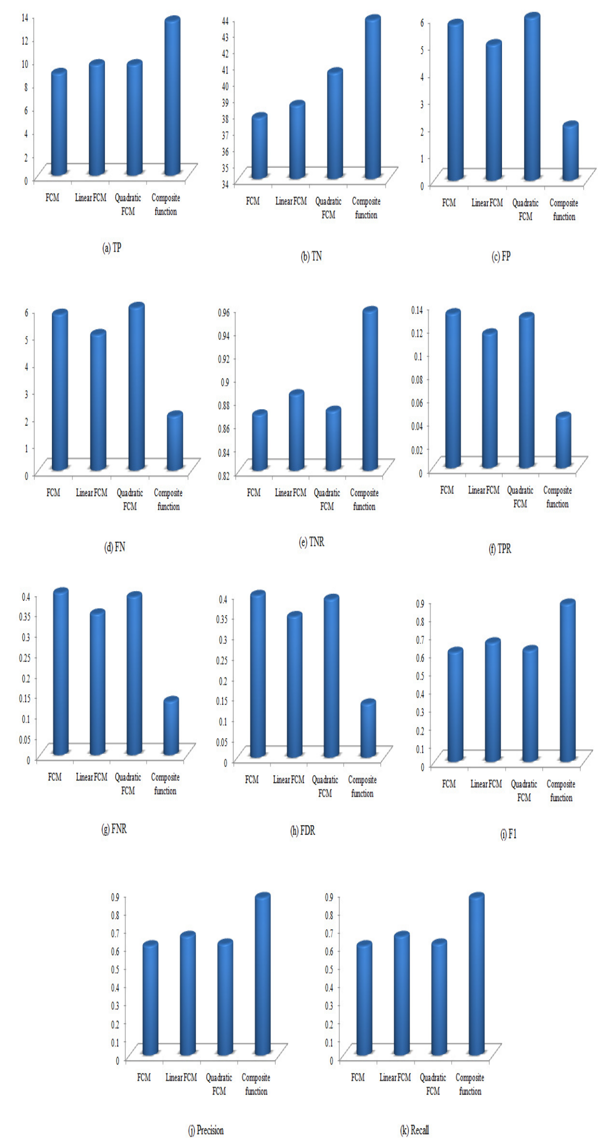

Figure.6 Average Value Attain From Performance Measure For Different Techniques By Applying Four Biomedical Query Articles

The average for True Negative Rate (TNR) FCM reveal 86%, Linear FCM reveals 88%, Quadratic FCM reveal 87% and Composite function reveal 95%. The average for True Positive Rate (TPR) FCM reveal 13%, Linear FCM reveals 11%, Quadratic FCM reveal 12% and Composite function reveal 4.3%. The average for False Negative Rate (FNR) FCM reveal 39%, Linear FCM reveals 34%, Quadratic FCM reveal 38% and Composite function reveal 13%. The average

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

for False Discovery Rate (FDR) FCM reveal 39%, Linear FCM reveals 34%, Quadratic FCM reveal 38% and Composite function reveal 13%. The average for F1 Score FCM reveal 60%, Linear FCM reveal 65%, Quadratic FCM reveal 61% and Composite function reveal 86%. The average for Precision FCM reveal 60%, Linear FCM reveal 65%, Quadratic FCM reveal 61% and Composite function reveal 86%. Finally, the average for Recall FCM reveal 60%, Linear FCM reveal 65%, Quadratic FCM reveal 61% and Composite function reveal 86%.

5. CONCLUSION

From the result, the biomedical literature is retrieved by using the preprocessing method, retrieved techniques, and clustering techniques. The Multi-Kernel Fuzzy c Means (MK-FCM) technique is utilized for clustering in this detached kernel functions such as FCM, Linear FCM, Quadratic FCM and composite kernel function are analyzed for several measuring metrics. The biomedical query article databases utilize in this work are a brain tumor, breast cancer, kidney stone and neovascularization. This work deals with different manipulation in result analysis, which shows that the proposed composite function (hybrid and quadratic) of FCM reveals superior performance in all forms of analysis eleven performance measures are taken in this process for analyzing four cases of validation such as single input, double input, triple input and four inputs, except single input remaining all cases the results are evaluated by eleven measures. Compared with other literature papers the results perform better and the classification rate is high. This proposed algorithm was executed on the working platform of MATLAB-16a. In future, other diverse procedures are joined to enhance the biomedical literature.

REFERENCES

[1] Andrea Addis, Giuliano Armano and Eloisa Vargiu,"Multiagent systems and information retrieval our experience with X.MAS",*Journal of Expert Systems with Applications*, Vol.39, pp.2509-2523, 2012.

[2] Pathak and Joshi,"Itransed Marathi Literature Retrieval Using SMS Based Natural Language Query",*Journal of Advances in Computational Research*, Vol.4, pp.125-129, 2012.

[3] Dina Vishnyakova, Emilie Pasche, Christian LoviS and Patrick Ruch,"Utilization of Ontology Look-up Services in Information Retrieval for Biomedical Literature",*Journal of Health technology and information*, Vol.186, pp.155-159, 2013.

[4] Faustina Johnson and Santosh Kumar Gupta,"Web Content Mining Techniques: A Survey",*International Journal of Computer Applications*, Vol.47, No.11, pp.44-50, 2012.

[5] Jinxian Lin and Shuangyang Zheng,"An optimizing search based on Kernel-based Fuzzy C-Means Clustering", *Journal of IEEE*, pp.1-3, 2009.

[6] Hanumanthappa and Deepa Nagalavi,"Identification and Extraction of Different Objects and its Location from a Pdf File Using Efficient Information Retrieval Tools",*Journal of IEEE*, pp.1-6, 2015.

[7] Roberto Perez-Rodríguez, Luis Anido-Rifona, Miguel Gomez-Carballaa and Marcos Mourino-Garcíaa,"Architecture of a Concept-Based Information Retrieval System for Educational Resources",*Journal of Sci. Comput. Program*, pp.1-39, 2016.

[8] David Perez-Rey, Ana Jimenez-Castellanos, Miguel Garcia-Remesal, Jose Crespo and Victor Maojo,"CDAPubMed: a browser extension to retrieve EHR-based biomedical literature",*Journal of BMC Medical Informatics and Decision Making*, Vol.12, No.29, pp.1-10, 2012.

[9] Julio Cesar dos Reis, Edemar Mendes Perciani and Rodrigo Bonacin,"Intention-based Information Retrieval of Electronic Health Records",*Journal of IEEE*, pp.217-222, 2016.

[10] Yanqing Jil, Hao Ying, John Tran, Peter Dews and Michael Massanari,"Integrating Association Mining into Relevance Feedback for Biomedical Literature Search",*Journal of IEEE*, pp.531-536, 2015.

[11] Anália Lourenço, Rafael Carreira, Daniel Glez-Pena, Jose Méndez, Sonia Carneiro, Luis Rocha, Fernando Díaz, Eugenio Ferreira, Isabel Rocha, Florentino Fdez-Riverola and Miguel Rocha,"BioDR: Semantic indexing networks for biomedical document retrieval",*Journal of Expert Systems with Applications*, Vol.37, pp.3444-3453, 2010.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[12] Yogan Jaya Kumar, Naomie Salim, Albaraa Abuobieda and Ameer Tawfik Albaham,"Multi document summarization based on news components using fuzzy cross-document relations", *Journal of Applied Soft Computing*, Vol.21, pp.265-279, 2014.

[13] Nookala Venu and Anuradha,"Two Different Multi-kernels Integration with Spatial Information in Fuzzy C-means Algorithm for Medical Image Segmentation",*Journal of IEEE,* pp.20-25, 2015.

[14] Kannan, Ramathilagam, Sathya and Pandiyarajan,"Effective fuzzy c-means based kernel function in segmenting medical images",*Journal of Computers in Biology and Medicine*, Vol.40, pp.572-579, 2010.

[15] Park, Jang, Kim, Huh and Chung,"The Design of Fuzzy C-Means Clustering based Neural Networks for Emotion Classification",*Journal of IEEE*, pp.413-417, 2013.

[16] Van Nha Pham, Long Thanh Ngo and Witold Pedrycz,"Interval-valued fuzzy set approach to fuzzy co-clustering for data classification",*Journal of Knowle dge-Base d Systems*, pp.1-13, 2016.

[17] Rafal Lancucki and Andrzej Polanski,"Automatic PDF Files Based Information Retrieval System with Section Selection and Key Terms Aggregation Rules"*Journal of Springer International Publishing* Switzerland, pp.251-258, 2016.

[18] Stephen Fitchett and Andy Cockburn,"An empirical characterization of file retrieval",*Journal of Human-Computer Studies*, Vol.74, pp.1-13, 2015.

[19] Poonam Yadav,"Document-Document similarity matrix and Multiple-Kernel Fuzzy C-Means Algorithm based web document clustering for information retrieval",*International Journal of Advanced Research in Computer and Communication Engineering*, Vol.3, pp.8317-8321, 2014.

[20] Paavo Nieminen, Ilkka Pölönen and Tuomo Sipola,"Research literature clustering using diffusion maps",*Journal of Informetrics*, Vol.7, pp.874-886, 2013.

[21] Hsin-Chien Huang, Yung-Yu Chuang and Chu-Song Chen,"Multiple Kernel Fuzzy Clustering",*Journal of IEEE Transactions on Fuzzy Systems*, pp.1-15, 2012.