Modified Cosine Similarity Measure based Data Classification in Data Mining

D. Mabuni

Abstract: Text data analytics became an integral part of World Wide Web data management and Internet based applications rapidly growing all over the world. E-commerce applications are growing exponentially in the business field and the competitors in the E-commerce are gradually increasing many machine learning techniques for predicting business related operations with the aim of increasing the product sales to the greater extent. Usage of similarity measures is inevitable in modern day to day real applications. Cosine similarity plays a dominant role in text data mining applications such as text classification, clustering, querying, and searching and so on. A modified clustering based cosine similarity measure called MCS is proposed in this paper for data classification. The proposed method is experimentally verified by employing many UCI machine learning datasets involving categorical attributes. The proposed method is superior in producing more accurate classification results in majority of experiments conducted on the UCI machine learning datasets.

Keywords: Clustering based similarity measure, Modified cosine similarity measure, Text classification, Text clustering.

I. INTRODUCTION

Similarity measures are particularly important in pattern recognition and computer vision. Online visual tracking based applications are increasing rapidly in computer vision and some of its applications are – traffic control, motion analysis, activity recognition, and video surveillance. Cosine similarity plays a vital role in the process of obtaining accurate finger printings. Text similarity measurements are frequently used for finding common features among the text documents in applications like information retrieval or query execution statements, text mining problems and information extraction.

Previous experiences have shown that the cosine similarity measure is the best method for text comparisons and text related data processing applications. Cosine similarity can be used to determine how similar two words or two sentences or two pages or two documents and it can also be used for Sentiment Analysis, Text Comparison, and image comparison and so on. Similarity function is considered as a building block for finding solutions for many of the real world applications particularly in data classification, clustering, search, and recommendations etc. Measure of similarity can be either qualitative or quantitative or both qualitative and quantitative. Quantitative similarity of a document is a numerical measure derived from the length of the document, number of key words, distinct words, and common words and so on.

Finding similarity between two documents is quintessential way of text mining applications and it is the extensively used method in machine learning and data mining. Finding documents with high similarity is really a challenging task in very large collections of documents. Vector space model is one convenient way for representing a set of documents. Cosine similarity measure is a multipurpose, multidimensional, and useful in multitude applications. Euclidean distance metric is generally not a suitable metric for high-dimensional data mining applications. Cosine similarity is one of the most popular text similarity finding measures.

Cosine similarity measure value between two sentences can be calculated as a normal dot product of their vector representation. In the case of desired information retrieval, the cosine similarity numerical value of two documents will range from 0 to 1. Note that cosine similarity measure is used to measure the cosine angle of two website’s content whether the website’s content is similar or not similar and it is how much in quantity.

Sometimes it may be necessary to find overall relationship among all the documents instead of finding relationship between two documents only. Dimensionality reduction is one such technique used for finding overall relationship among all the attributes or documents. Day by day large number of documents is being generated in the Internet. Better organization of all these documents is needed in a way that will be very easy for the end users to access and use the desired documents conveniently in less time.

In order to characterize the similarity between two images cosine similarity measure can be applied in different image processing applications, graphics applications, and mechanical objects designing such as face recognition, pattern recognition, and template recognition, aircraft designing. In general, the cosine similarity allows comparing of two objects, customers, files, persons, companies, colleges, universities, shops, product profiles, villages, cities, mechanical objects, electrical objects and in particular optimal management of text documents etc. Cosine similarity finding formula is very useful to find solutions to the many real world problems ranging from small complexity to high complex tasks such as

1) Finding the most similar document to the given document
2) Finding top ten of the most similar documents for the given document
3) Are there any set of customers who are likely to purchase a specific product.
4) Is there any rule for clustering or grouping the products
5) Are there any recommendations to increase the sales of a
specific product based on the customer’s purchases.

In order to improve quality of the similarity match a set of
data transformations are required. These transformations may
be simple data conversions such as converting given data into
normalized data or a set of matrix operations etc. Normalization plays an important role in achieving understandable, comparable, useful and convenient similarity
finding results. Also in data mining standard formulas are
already available for data normalization. In order to compare
one short document with one long document the better way is
first normalize the documents and then compare them.
Normalization helps ease of computing accurate similarity
measure results.

II. LITERATURE SURVEY

Ainura Madlylova and Sule Gunduz Oguducu [1] proposed
a new cosine similarity based algorithm for finding semantic
similarity between documents and experimentally verified
that the proposed method is faster than all the existing
methods. Cihan Varol and Sairam Hari [2] have proposed a
new algorithm for finding near duplicate text documents. C.
Liu [3] has proposed two similarity measures whitened cosine
similarity measure and with-in class whitened Cosine
Similarity Measure and the effectiveness of these two
measures is verified in face recognition experiments. Delphine Charlet and Geraldine Damnati [4] proposed question-question similarity task for the purpose of
developing semantic textual similarity measures.

Dong Wang et. al. [5] has proposed a weighted local cosine
similarity for visual tracking. H. Mohana and M. Suriakala
[6] proposed a new Tuned cosine similarity measure for
prediction because the normal similarity measures like
Cosine, Pearson correlation coefficient (PCC), and Jaccard
are not effective for prediction and the experimental results
have shown that Tuned cosine measure outperforms over all
the previous similarity comparison measures in terms of
accuracy in prediction. K.M. Hammouda and M.S. Kamel [7]
two ideas are proposed for document clustering and the first
one is called phrase-based document index model and the
second is called incremental document clustering algorithm
based on the maximizing the tightness of the clusters. M.M.
Zanjireh [8] has proposed text similarity finding measure
based on term weights and the minimum number of terms
appeared in the document. Sahar Sohangir and Dingding
Wang [9] have proposed a new measure for text similarity
finding and experimentally verified that the proposed method
is effective when compared with existing text similarity
finding measures.

S. Sohangir and D. Wang [10] Cosine similarity based on
the standard Euclidean distance value is currently considered
to be one of the most widely used similarity measurements but
Euclidean distance metric is not effective and correctly
suitable metric for managing probabilities that are commonly
used in text data analytics. Authors proposed and applied a
new similarity measure for documents management based on
sqrt-cosine similarity. Some examples for document

similarity are document classification, document clustering,
document search, document summarization, and query-based
search etc. Document similarity is commonly used in natural
language processing. Now a days large number of text
documents are flowing over the Internet. This has brought
about a special need for efficient, effective and automatic
management of documents in areas such as classification,
clustering and other related tasks. Shuai Han et. al. [11]
proposed new cosine similarity measure based technique to
improve positioning performance of fingerprinting.

T. Chanwimalueang and D. P. Mandic [12] have proposed
two entropy estimators based on the similarity of embedding
vectors, the first one is called cosine similarity entropy and the
second one is called multi scale cosine similarity and shown
that both are superior to normal entropy. Venkatakaneshbabu
k et. al. [13] proposed new similarity measure and it is named
Maxwell–Boltzmann Similarity Measure (MBSM) and experimentally verified that the proposed measure
outperforms all the existing measures in similarity finding
applications.

III. COSINE SIMILARITY MEASURE FOR DATA
CLASSIFICATION OR CLUSTERING

Cosine similarity is a similarity measure and it is defined as
the cosine of the angle between two selected n-dimensional
vectors in the n-dimensional space. It is equal to the dot
product of the two vectors divided by their product of
magnitudes. This measure is equal to the similarity measure
between two non-zero vectors of an inner product space. The
fundamental goal of this similarity measure is to find the
similarity quantitatively. Soft cosine metrics are far better for
finding similarity in data clustering and classification
applications.

Cosine similarity is a standard metric and it is generally
used to find the similarity between two documents without
considering their size. Cosine similarity increases as the angle
decreases. The standard cosine similarity metric formula can
be interpreted and adjusted conveniently for different
applications according to the specifications.

\[
\text{Cosine Similarity} = \cos(\theta) = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}}
\]

Where \(A_i\) and \(B_i\) are components of vectors \(\vec{A}\) and \(\vec{B}\) and

\[
\|\vec{A}\| = \sqrt{A_1^2 + A_2^2 + \cdots + A_n^2}
\]

\[
\|\vec{B}\| = \sqrt{B_1^2 + B_2^2 + \cdots + B_n^2}
\]

Cosine Distance=1– Cosine Similarity

IV. PROPOSED MODIFIED COSINE SIMILARITY
(MCS) MEASURE FOR DATA CLASSIFICATION

In the training dataset class counts of distinct categorical
values of each attribute is represented in the form of modified
special vectors.
For example, if the attribute color has four distinct categorical values then class counts of each categorical value is represented conveniently. Modified cosine similarity measure is computed by dividing the product of class counts with normalized lengths of categorical values across all the distinct values of each categorical attribute. The modified cosine similarity formula is used for finding similarities in the sample training dataset.

A sample training dataset is used in this paper for explaining the modified cosine similarity computations and experimental results. This dataset contains four predictor attributes and one class attribute with two class labels 0 and 1. Class label 1 means institute running permission is given and the class label 0 means permission is not given.

### TABLE-1 Sample Training Dataset

| Institute | Affiliation | Urban | Facilities | Permission |
|-----------|-------------|-------|------------|------------|
| School    | Three       | No    | Bad        | 0          |
| School    | Three       | No    | Good       | 1          |
| Inter     | Two         | No    | Bad        | 0          |
| Degree    | One         | Yes   | Bad        | 0          |
| Degree    | One         | Yes   | Good       | 1          |
| School    | Two         | No    | Bad        | 0          |
| School    | One         | Yes   | Bad        | 1          |
| Degree    | Two         | Yes   | Bad        | 1          |
| Degree    | Two         | No    | Good       | 1          |
| Inter     | Two         | Yes   | Good       | 1          |
| Inter     | Three       | Yes   | Bad        | 1          |
| Degree    | Two         | Yes   | Good       | 1          |
| Degree    | Two         | No    | Good       | 0          |

Institute attribute has three distinct categorical values, School, Intermediate College, and Degree College. Institute attribute represents type of institute currently running. Affiliation attribute has three distinct categorical values One, Two, and Three. Affiliation will be given to the concerned institute for one year, or two years or three years. Urban attribute has two distinct categorical values, Yes and No. Yes means institute is established in urban area and No means institute is established other than urban area. Facilities attribute represents facilities available in the current institute and all the facilities are grouped together and combined values are called Good and Bad. Good means more than enough facilities are available and Bad means facilities are not enough in the current institute.

Government authority has to supervise and give permission for running the institute based on all the attribute values of a particular institute. Once all the specified requirements are satisfactory then the permission will be given or permission will not be given by the concerned government authority.

The formula representation of modified cosine similarity measure is

\[
MCS(\text{Attribute}) = \frac{\text{sum of products of class counts of the attribute}}{\text{product of normalization factors}}
\]

Institute attribute has 3 distinct categorical values accordingly 3 groups of class counts are created and then modified cosine similarity is applied for finding similarity among all these groups with respect to one attribute of the training dataset.

### TABLE-2 class counts of Institute attribute

| Institute | Permission-1 | Permission-0 |
|-----------|--------------|--------------|
| School    | 2            | 3            |
| Inter     | 4            | 0            |
| Degree    | 3            | 2            |

### TABLE-3 class counts of Affiliation attribute

| Affiliation | Permission-1 | Permission-0 |
|-------------|--------------|--------------|
| One         | 3            | 1            |
| Two         | 4            | 2            |
| Three       | 2            | 2            |

### TABLE-4 class counts of Urban attribute

| Urban | Permission-1 | Permission-0 |
|-------|--------------|--------------|
| Yes   | 6            | 1            |
| No    | 3            | 4            |

### TABLE-5 class counts of Facilities attribute

| Facilities | Permission-1 | Permission-0 |
|------------|--------------|--------------|
| Good       | 3            | 3            |
| Bad        | 6            | 2            |

\[
MCS(\text{Institute}) = \frac{2 \cdot 3 + 4 \cdot 0 + 3 \cdot 2}{\sqrt{2^2 + 4^2 + 3^2 + 2^2}} = \frac{12}{\sqrt{29}}
\]

\[
MCS(\text{Affiliation}) = \frac{3 \cdot 1 + 4 \cdot 2 + 2 \cdot 2}{\sqrt{3^2 + 4^2 + 2^2}} = \frac{15}{\sqrt{29}}
\]

\[
MCS(\text{Urban}) = \frac{6 \cdot 1 + 3 \cdot 4}{\sqrt{6^2 + 3^2 + 2^2}} = \frac{18}{\sqrt{45}}
\]

\[
MCS(\text{Facilities}) = \frac{3 \cdot 3 + 6 \cdot 2}{\sqrt{3^2 + 6^2 + 3^2 + 2^2}} = \frac{18}{27.6586} = 0.6508
\]
Modified Cosine Similarity Measure based Data Classification in Data Mining

\[
MCS(\text{Facilities}) = \frac{9 + 12}{\sqrt{9 + 36\sqrt{9}+4}} = \frac{21}{\sqrt{45}\sqrt{13}}
\]

\[
MCS(\text{Facilities}) = \frac{21}{6.7062 \times 3.606} = \frac{21}{24.1898} = 0.8681
\]

\[
MCS(\text{Facilities}) = \frac{1 + 2}{\sqrt{1 + 1\sqrt{1}+4}} = \frac{3}{\sqrt{10}}
\]

\[
MCS(\text{Facilities}) = \frac{3}{3.1622} = 0.9487
\]

| Attribute Name | MCS value |
|----------------|-----------|
| Institute      | 0.6179    |
| Affiliation    | 0.9285    |
| Urban          | 0.6508    |
| Facilities     | 0.8681    |

**TABLE-6**

In TABLE-6 MCS value of Institute attribute is minimum and therefore the best data classification attribute is Institute. So, data are classified based on distinct categorical values of the attribute Institute.

Same procedure is used for sub divided data groups.

**TABLE-7 Institute = “School” sub data**

| Affiliation | Urban | Facilities | Permission |
|-------------|-------|------------|------------|
| Three       | No    | Bad        | 0          |
| Three       | No    | Good       | 0          |
| Two         | No    | Bad        | 0          |
| One         | Yes   | Bad        | 1          |
| Two         | Yes   | Good       | 1          |

**TABLE-8 class counts of Affiliation attribute**

| Affiliation | Permission-1 | Permission-0 |
|-------------|--------------|--------------|
| One         | 1            | 0            |
| Two         | 1            | 1            |
| Three       | 0            | 2            |

| Affiliation | Permission-1 | Permission-0 |
|-------------|--------------|--------------|
| One         | 1            | 0            |
| Two         | 1            | 1            |
| Three       | 0            | 2            |

**TABLE-9 class counts of Urban attribute**

| Urban | Permission-1 | Permission-0 |
|-------|--------------|--------------|
| Yes   | 2            | 0            |
| No    | 0            | 3            |

\[
MCS(\text{Urban}) = \frac{2 \times 0 + 0 \times 3}{\sqrt{2^2 + 0^2 \sqrt{2^2} + 3^2}} = \frac{0}{\sqrt{2^2 + 0^2 \sqrt{2^2} + 3^2}} = 0
\]

\[
MCS(\text{Urban}) = \frac{0}{4 + 0/0 + 9} = \frac{0}{6} = 0
\]

**TABLE-10 class counts of Facilities attribute**

| Facilities | Permission-1 | Permission-0 |
|------------|--------------|--------------|
| Good       | 1            | 1            |
| Bad        | 1            | 2            |

\[
MCS(\text{Facilities}) = \frac{1 \times 1 + 1 \times 2}{\sqrt{1^2 + 1^2 \sqrt{1^2} + 2^2}}
\]

**TABLE-11 Similarity measure values**

| Attribute Name | MCS value |
|----------------|-----------|
| Affiliation    | 0.31624   |
| Urban          | 0         |
| Facilities     | 0.9487    |

In TABLE-11 Urban attribute is the best data classification Attribute because its MCS score is minimum. So, data in TABLE-7 is classified based on distinct categorical values of Urban attribute. The resulted data classification tables are shown in TABLE-12 and TABLE-13 respectively.

**TABLE-12 Urban = “No” sub data**

| Affiliation | Facilities | Permission |
|-------------|------------|------------|
| Three       | Bad        | 0          |
| Three       | Good       | 0          |
| Two         | Bad        | 0          |

**TABLE-13 Urban = “Yes” sub data**

| Affiliation | Facilities | Permission |
|-------------|------------|------------|
| One         | Bad        | 1          |
| Two         | Good       | 1          |

**TABLE-14 Institute = “Inter” sub data**

| Affiliation | Urban | Facilities | Permission |
|-------------|-------|------------|------------|
| Three       | No    | Bad        | 1          |
| One         | Yes   | Bad        | 1          |
| Two         | No    | Good       | 1          |
| Three       | Yes   | Bad        | 1          |

Data in the TABLE-14 is pure and hence further data classification is not required.

**TABLE-15 Institute = “Degree” sub data**

| Affiliation | Urban | Facilities | Permission |
|-------------|-------|------------|------------|
| Two         | No    | Bad        | 1          |
| One         | Yes   | Bad        | 1          |
| One         | Yes   | Good       | 0          |
| Two         | Yes   | Bad        | 1          |
| Two         | No    | Good       | 0          |

**TABLE-16 class counts of Affiliation attribute**

| Affiliation | Permission-1 | Permission-0 |
|-------------|--------------|--------------|
| One         | 1            | 1            |
| Two         | 2            | 1            |
\[
MCS(\text{Affiliation}) = \frac{1 + 1 + 2 + 1}{\sqrt{1^2 + 2^2 + \sqrt{1^2 + 1^2}}}
\]
\[
MCS(\text{Urban}) = \frac{1 + 2}{\sqrt{1 + 4 + \sqrt{1 + 1}}} = \frac{1}{\sqrt{5}}
\]
\[
MCS(\text{Urban}) = \frac{1}{\sqrt{10}} \times \frac{1}{3.1622} = 0.31624
\]

**Table 17** class counts of Urban attribute

| Urban | Permission-1 | Permission-0 |
|-------|--------------|--------------|
| Yes   | 1            | 1            |
| No    | 2            | 1            |

**Table 18** class counts of Facilities attribute

| Facilities | Permission-1 | Permission-0 |
|------------|--------------|--------------|
| Good       | 0            | 2            |
| Bad        | 3            | 0            |

**Table 19** Similarity measures

| Attribute | MCS value |
|-----------|-----------|
| Affiliation | 0.31624 |
| Urban      | 0.31624 |
| Facilities | 0        |

In Table 19 Facilities attribute is the best data classification attribute because its MCS score is minimum. So, data in Table-15 is classified based on distinct categorical values of Facilities attribute. The resulted data classification tables are shown in Table-20 and Table-21 respectively.

**Table 20** Facilities = “Good” sub data

| Affiliation | Urban | Permission |
|-------------|-------|------------|
| One         | Yes   | 0          |
| Two         | No    | 0          |

**Table 21** Facilities = “Bad” sub data

| Affiliation | Urban | Permission |
|-------------|-------|------------|
| Two         | No    | 1          |
| One         | Yes   | 1          |
| Two         | Yes   | 1          |

Table 20 and Table 21 are not further classified because they are pure.

V. ALGORITHM

Algorithm Modified-Cosine-Similarity (D)

**Input**

D is the input dataset

**Output**

The best attribute for data classification

1. A = the set of attributes in the training dataset, D
2. for each attribute i in the set A do
3. CAT=set of distinct categorical values of the attribute i
4. find groups of class counts for each value in CAT
5. find modified cosine similarity for each attribute i
6. array[i] = cosine similarity measure for the attribute i
7. end for i loop
8. minimum = 999999
9. location = 0
10. for i = 1 to length of array[] do
11. if(array[i] < minimum )
12. minimum = array[i]
13. Location = i
14. end if
15. end for i loop
16. A_i = the best data classification attribute
17. return A_i

The proposed algorithm called modified cosine similarity (MCS) is an effective method for data classification. It computes a numerical measure value between 0 and 1. In the extreme case the value 0 measures the best data classification and the value 1 measures the worst data classification. Value 0 is more desirable in many applications. Modified cosine similarity algorithm is implemented as a function which finds and then returns the best attribute for data classification in the main algorithm.

VI. RESULTS

**Table 22** Experimental results for the UCI machine learning datasets

| Dataset Name | Training dataset size | Test dataset size | C4.5 accuracy | MCS accuracy |
|--------------|-----------------------|-------------------|---------------|-------------|
| Nursery      | 9719                  | 3240              | 86.3580       | 86.3580     |
| Breast Cancer| 192                   | 95                | 61.0526       | 75.7894     |
| Car data     | 1296                  | 432               | 81.25         | 77.0833     |
| Balance scale| 469                   | 156               | 45.5128       | 60.2564     |
| Primary tumor| 254                   | 85                | 55.2941       | 75.2941     |
| Hayes Roth   | 132                   | 28                | 50.0          | 50.0        |
| SPEC         | 80                    | 187               | 58.8235       | 72.7272     |
| Lymphography | 112                   | 36                | 27.7777       | 41.6666     |
| permission   | 14                    | 14                | 100.0         | 100.0       |
Modified Cosine Similarity Measure based Data Classification in Data Mining

First bar (blue color bar) represents C4.5 accuracy and the second bar (red color bar) represents proposed algorithm (MCS) accuracy. The accuracy of the proposed method is superior in all the cases of the employed UCI machine learning repository training datasets except in the case of car dataset. Labels are given only for the proposed method but not for the existing method.

VII. CONCLUSION

Cosine similarity is popularly used in many text processing applications such as documents classification, clustering, querying, and information retrieval etc. A new clustering based modified Cosine similarity measure is proposed in this paper for data classification. Proposed similarity measure is applied on UCI machine learning datasets. The results of the experiments revealed that the proposed similarity measure outperforms the many of the existing similarity measures. In the future different types of modified cosine similarity measures will be discussed to the greater extent for effective, accurate, general and efficient data classification with normalization measures concept.

REFERENCES

1. Ainura Madylova ; Sule Gunduz Oguducu, “A taxonomy based semantic similarity of documents using the cosine measure”, Published in: 2009 24th International Symposium on Computer and Information Sciences, Date of Conference: 14-16 Sept. 2009, Date Added to IEEE Xplore: 23 October 2009.
2. Cihan Varol and Sairam Hari, “Detecting near-duplicate text documents with a hybrid approach”, sage journals, journal of information science, First Published March 24, 2015 Research Articlehttps://doi.org/10.1177/0165551515577912.
3. C. Liu, “The Bayes Decision Rule Induced Similarity Measures”, IEEE Transactions on Pattern Analysis and machine Intelligence, Volume: 29, Issue 6, June 2007, pages: 1086 – 1090.
4. Delphine Charlet and Geraldine Dannati, “Soft-Cosine Semantic Similarity between Questions for Community Question Answering” 'Orange Labs Lannion, France, Proceedings of the 11th International Workshop on Semantic Evaluations (SemEval-2017), pages 315–319, Vancouver, Canada, August 3 - 4, 2017, e 2017.
5. Dong Wang , Huchuan Lu and Chunjuan Bo, “Visual Tracking via Weighted Local Cosine Similarity”, Published in: IEEE Transactions on Cybernetics (Volume: 45, Issue:9,Sept.2015),Page(s): 1838 – 1850.
6. H. Mohana and M. Suriakala, “Tuned Cosine Similarity Measure To Alleviate Data Sparsity Issues For Personalized Recommendation”.

Published in: 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS) Date of Conference: 20-22 Dec. 2018, Date Added to IEEE Xplore: 25 July 2019.
7. K.M. Hammouda and M.S. Kamel, “Efficient phrase-based document indexing for Web document clustering”, Published in: IEEE Transactions on Knowledge and Data Engineering (Volume: 16, Issue: 10 , Oct. 2004 ), Page(s): 1279 – 1296, Date of Publication: 24 August 2004.
8. M.M. Zanjireh, “Pairwise document similarity measure based on present term set”, Springer, Journal of Big Data volume 5, Article number: 52 (2018).
9. Sahar Sohangir and Dingding Wang, “Document Understanding Using Improved Sqrt-Cosine Similarity”, Published in: 2017 IEEE 11th International Conference on Semantic Computing (ICSC), Date of Conference: 30 Jan.-1 Feb. 2017, Date Added to IEEE Xplore: 30 March 2017.
10. S. Sohangir and D. Wang, “Improved sqrt-cosine similarity measurement”, Springer, Journal of Big Data, volume 4, Article number: 25 (2017).
11. Shuai Han, Cong Zhao, Weixiao Meng and Cheng Li, “Cosine similarity based fingerprinting algorithm in WLAN indoor positioning against device diversity”, Published in: 2015 IEEE International Conference on Communications (ICC), Date of Conference: 8-12 June 2015, Date Added to IEEE Xplore: 10 September 2015.
12. T. Chanwimalueang and D. P. Mandic, “Cosine Similarity Entropy: Self-Correlation-Based Complexity Analysis of Dynamical Systems”, Department of Electrical and Electronic Engineering, Imperial College, SW7 2AZ London, UK.
13. Venkatanareeshbabu k, Mainak B. Damodar ReddyE, K.J.Ravi Prasad, Ijasit S. Suri, “A Mechanics-Based Similarity Measure for Text Classification in Machine Learning Paradigm”, Published in: IEEE Transactions on Emerging Topics in Computational Intelligence (Volume: 4, Issue: 2, April 2020 ), Page(s): 180 – 200 Date of Publication: 24 August 2018.

AUTHORS PROFILE

D. Mabuni, completed M.Sc. (Computer Science), MCA and M.Phil. (Computer Science). Currently working as Assistant Professor in the Department of Computer Science at Dravidian University, Kuppan, Andhra Pradesh, India. My interested research areas are Data Mining, Databases, and User Interfaces.