

GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES
LSSFCA: LATENT SEMANTIC SUBSET FEATURE CLUSTER ANALYSIS FOR
DEVELOPING EDUCATIONAL SYSTEM USING ATTRIBUTE BASED CONTENT
SIMILARITY MEASURE IN DATA WAREHOUSING

R. Sundara Bharathi^{*1} & Dr. R. Mala²

^{*1}Research Scholar, Bharathidasan University, Marudupandiyar College, Thanjavur, India

²Assistant Professor, Department of Computer Science, Alagappa University Model Constituent College of Arts & Science, Paramakudi, India

ABSTRACT

Nowadays educational development is great approach for various fields to improving the knowledge process for minimizing the work process. To deal the educational facts problems in web mining resource become problematic due to improper information accessing to the students due from online learning progress. This fact is occurred due to extraction of improper information obtained because of failed techniques leads time complexity creates burden. To develop a semantic approach using knowledge process that improve the performance. Several implementation process the semantic measure using cluster algorithm for predicting results from educational documents. In this paper, we propose a Latent semantic subset feature cluster analysis (LSSFCA) for developing educational system using attribute based content similarity measure and subset feature clustering algorithm is implemented. This progress the educational documents with relevance identification of search key terms to obtain the similarity measure. The sentence level semantic approach identifies the similarity features based key terms and cluster the document by level to provide the information to the students. This method produce high efficiency to improve the educational learning resources, to solve the time of evaluation approach intents best evaluation results.

Keywords: semantic similarity, cluster evaluation, latent indexing, educational resource, sentence level subset features.

I. INTRODUCTION

The educational services provide information to the user through on learning progress to access the information for various knowledge learning course, especially called online course. This services are provides from educational institution through web resource's in online. The students are studied the though accessing the relational content analysis or searches to improve the capability by getting the fact of educational documents. They are various service like webinars, e-learning, dashboard to provide the services in different ways to improve educational process. The web miming is the great invention to extract the knowledge process based on the semantic relational search from education web documents.

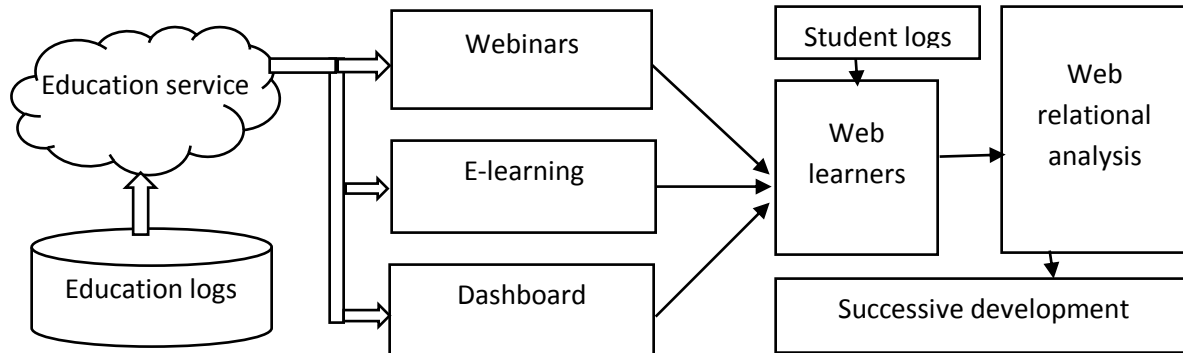


Figure 1. Educational data mining in warehousing

Educational development fact on new techniques contains the content analysis and deliver to the learners on different ways as shown in figure 1. The ware housing contains various educational resources on web through attaining the subjectivity services by analyzing semantic similarity measure. By utilizing the content analysis and extracting feature have the subjectivity measure to cluster the educational documents. By analyzing the semantic similarity measure using latent terms by extracting the features for the student query. The query evaluation process enhance the subjectivity measure of content evaluation by information similarity that are from educational web documents .the cluster resembles the group of key terms evaluation by suing selected optimal features. By the extraction of selected features summarization task of relational sentence to make similarity .In this only the key terms of relational similarity count is taken to reduce the frequent state to improve the cluster accuracy by retrieving the deliver content from wen educational resources.

Some sentences in the documents represent some aspects of their contents to some extent. Moreover, speed will be an important factor while incorporating the summarization facility on the web. So, extraction based summarization is still useful on the web. The extractive multi-document summarization can be concisely formulated as extracting important textual units from multiple related documents, removing redundancies and reordering the units to produce the fluent summary. Reducing the time relevancy approach of summarization using feature extraction by multi objective latent semantic keyword extraction to deal the subjectivity measurement.

In this research, we propose ansemantic relational to deliver better lectures and contents to the students considering in the remote areas, and hence to improve the nature of education and interest. Cluster evaluation resembles the Web technology as its fundamental technical infrastructure to deliver knowledge for its improvement. As the current trend of academic and modern realities is to increase the use of educational development, shortly a higher demand for technology bolster is expected. The subjectivity measure enhance the cluster group by evaluation measure. Specifically, educational apparatuses supporting the major undertaking of direction design ought to provide automated help for the investigation, design, documentation, implementation, and deployment of the guideline using Web learners.

II. LITERATURE SURVEY

There are many approaches has been discussed for the development of educational development system. We discuss few of them here in this section.

The educational development optimize various semantic relation to analyze the student capability or knowledge delivery for improving the educational environment .mostly web base learning is analyzed through data mining techniques to deliver content to the learners [1].To analyses the subjects from semantic search using summarization techniques [2]. This creates an extended text elaboration to group the documents with time complexity, also the specificity level is not depends on relational terms [3]. The content similarity measure holds the keyword count on maximal relativity measure to deliver the education contents [4]. This doesn't satisfy the learner to describe the

originality of behavioral based learning providence [5]. So the development approaches depending to the cluster and classification process.

By reducing the mismatch of data relevancy measurement the semantic procedure is change into classification by analyzing the semantic level keyword terms [6]. So the information extraction is the essential needs to analyze the predominant term of access to the relevancy. The depending nature of data mining the information related to another similar way to provide the subjectivity to the learner [7, 8].after developing summarization techniques define the related content analysis to the text evaluation to proceed the content analysis [9]. Automatic process defines the leading relational approach for personalized search to depend the student interest [10]. Also improving quality be analyzed through multimedia content analysis [12] .the text mining process patterns are relational subjectivity to extract the relational dependency by specific measure of subjects required by students capability of learning. ELearning produce the centralized approach to mine results from educational resource carried in web documents [13]. The learning subjects are directly through web based internet of approach by cloud providers [14]. The relational approach are carried to use observer the sentence relation and the selected service to the students [15]. The sentence level semantic cluster mostly group the relevance case content to rank the learning series from web resources. The implementation mostly group the content measure not basis on semantic relational measure [16]. This leads non mismatch data to analyze the education proprieties reduce the efficiency level in educational resources

Due to development of advanced learning techniques, the materials to be analyzed through knowledge learning process through data mining [17]. This provide the online learning projection with time complexity problems.to detect the learning style through the cluster behavioral analysis model and classification model [18, 19]. The semantic relation among the educational web documents are processed by the measuring the concurrence measure kills observed from students [20]. As the way the word or sentence level query is analyses from the student search content taken is group the subjects to deliver students.

III. IMPLEMENTATION OF PROPOSED SYSTEM

The development of educational resource using web mining analysis in the tremendous approaches for applying various techniques. Accessing the knowledge based relation content using semantic extraction on differential fields promotes the students to gain the learning capabilities. Same as the development educational resources provides the learning documents through attract the student's capability of service providence to improve the educational progress. Our idea integrates the new invention of latent semantic indexing to improve the information access based on the topic content analysis in web documents. The process of web mining resources resembles the relational analyses problem using subset features selection to extract the key term to support cluster evaluation. The feature selection reduce the time complexity level to improve the clustering efficiency.

By construction the queries from students on educational institutions became various problems due to effectiveness of cluster or semantic evaluation of information extraction. To propose a Latent semantic subset feature cluster analysis (LSSFCA) to select the sufficient redundant features to analyses the documents for developing educational system, also using selected attribute based content similarity measure scores for further applying the sentence level based clustering algorithm for optimize the cluster evaluation result. This standardized the educational document evaluation approach in the point of content search using semantic similarity measure. The proposed system impacts the content using latent semantic feature based on education development system using selected semantic measure to obtain efficient learning system, subset feature based sentence-clustering algorithm is intended to identify topic key terms of search content from students that are taken and selects the representative features in search sentences from the semantic similarity measure to form topic related feature based appropriate clusters to improve the optimization.

The semantic similarity measure that identifies the topic relational scores to intent selected features on reducing cluster probabilistic maintains the time complexity. Also his measure improve the deep discussion of content evaluation on related query search from educational web documents the relative features identifies maintermes of semantic content to indexing the specific points.

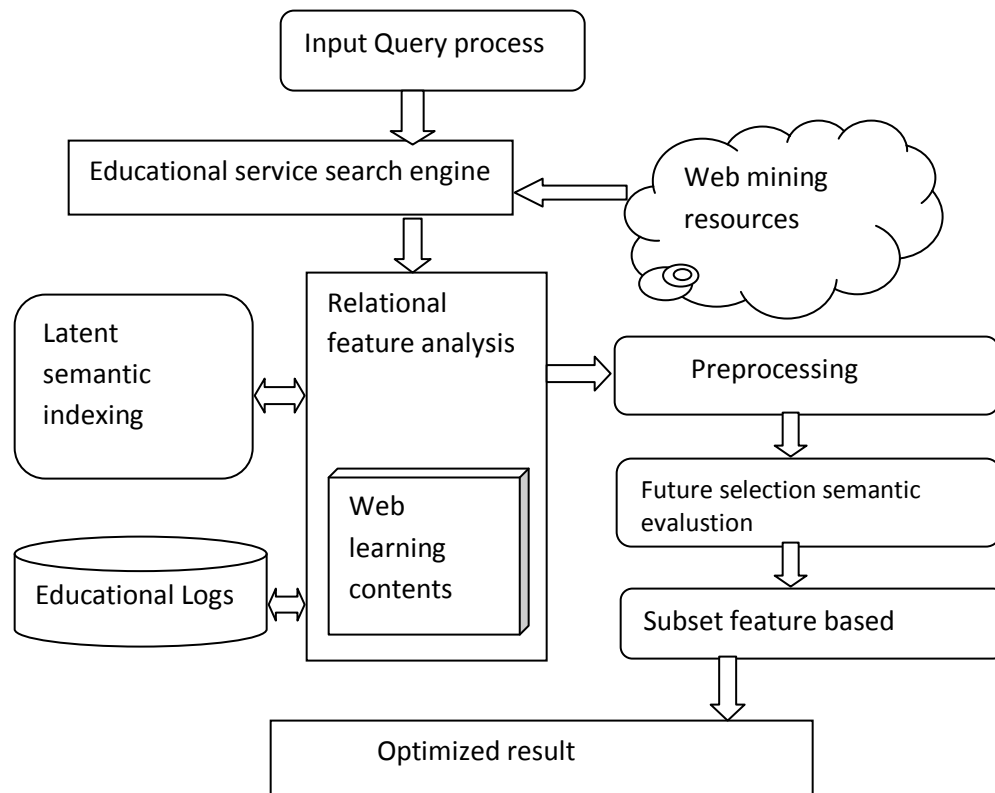


Figure 2: Architecture diagram Latent semantic subset feature cluster analysis

The querying process of content evaluation resembles the domain content analysis model to analyze the contents from educational resources. Figure 2 shows the architectural process of our proposed system. Mainly this consider the two factors, one is for query content analysis and another is relative analysis for specifying relational terms which is for ground truth values analyzed from semantic similarity content. Form this fact the relational features are computationally process to extract main terms are features. With the selected features the subset are evaluated to consider as index terms to search queries. Mainly the semantic relations to identify the match case relational features depending the topic similarity measure. Also the document similarity measure depends the linguistic topic detection variable to groups the relations among clusters set and features based clusters are optimized.

3.1 Preprocessing

The initialization stage begins the text parts which the selected documents are from educational resources. The preprocessing initialize the noise removal from the part of text with analysis the relational sentences. To remove the stop words and other non-related terms are depended to identify the relational part. Which the special symbols not in case sensitive depends to remove the frequent terms. These terms are dependent query analysis the tokenization as count words. This resembles the complex nature of content focus the main terms of latent terms

Complex nature of preprocessing enhance the extraction from the query inputs and educational documents set to prepare for nature of proposes semantic measure. The preprocessing includes: extraction of stop words, tokens represented terms, keywords holding, term dependencies content.

Input: educational web document set.

Output: poised noise removed count documents Ps

Start

Step 1 :Input dataset Elds → Ec1,Ec2,Ec3..are the records

```

For (Ec check NULL)
    Null fields of doc remove and Verify contents
    List to Reorder educational records
    ECs ← Ec(non emptyedu sets)
End for
For (remove distinctsedu sets)
    Fill Eds → content matches
End for
Step 2: Analyze index terms of content
    If (EICS != attribute contentindex As)
        Step 3: find all the distinct attributes EAs
            EAs =  $\int_{i=1}^{size(Cds)} \sum ECds(cintent\ index\ i). Attribute \ni EAs \rightarrow$  sentence case index terms
            Removing non content stop words for edu sets
            EWCs =  $\int_{i=1}^{size(Cds)} \sum EWCds(i). query \ni Cs < - - non\ related\ terms$ 
        Step 4: for (all doc == stop word remove EWdi).
            If EWdi contains original Ecs < - - no related index terms
                Add the count words to the process the Eps
            Else
                Compute all non-relational data signs and other unreadable as noisy contents
            End if
        End for
    End if
Stop

```

With the chosen term set, the method estimates removal of noise consideration with the result returned by different point of services. The stop words has been determined based on the occurrence of the terms of input term set and the list of conditions being selected from the retrieved documents from query evaluated links.

3.2 Active latent content mining.

Latent content learning analysis the semantic relations which taken the features from subset terms from query and educational set. This process depends the identification of query keyword terms from obtained relational terms as considered feature set. This identifies the related feature based on input query from the student. Feature vector points the integrated with relational points of count words. Count words have the weightage with identity point of query terms with related search point

Algorithm:

Input: preprocessed document set Pds.

Output: Feature Vector Fv.

Step 1: Start

 Compute the count word from Pds page

PTs = (total repeat query terms – count terms(PTs))

Step 3 Total count terms CountSet = $\sum \{Count\ set\} \gg webinar$

For (total term-hidden relation)

 Generate feature vector select case

 Point feature terms of fv = { serviced, Ts, ActSet}.

Step 4: select the features

For each domain EDi from So

 Information repeated to generate query Q → sm

Step 5: relational measure of count words

Sm → countered hidden relation Nr = $\sum feature \in Gi$

 End

Stop.

Feature vector can be relevant case on subjectivity by count words and its calculated using higher space vector model which can be features to select to according to the semantic similarity measure between user's request and candidate services

3.3 Relative semantic subjectivity measure.

The relative query identifies the co-relation among the count words produces a set of result for the input query which are taken form educational documents. The user submits a query and form state matrix produces a set of the relational word matches. The semantic relation measure represents the strength of relevancy the result has to the input query. However, there exist numerous documents for a query which has been categorized in any form. Validating the relevancy of the student query content towards the query is necessary. The semantic analysis the query relational by subjectivity measure of repeated terms.

The steps are given below shows that

Input: educationallearning docsELd, query list Q(i),subjectivity access(Si) output : subjectivity measure

Step 1: For each search ELd→subjectivity analytics Q(i)

 Compute the relational word co-occurrence

$$\text{Concurrence measure CSLFreq} = \frac{\sum Al(i).query == Si \text{ subjectivity access}}{\text{total squery accesse +dterm repeated query}}$$

Step 2 Compute the relational semantic subjectivity analysis.

$$\text{RSA} = \frac{\sum Al(i).query == Si \text{ subjectivity}}{\text{total doc accessed}} * Sl(\text{semtic subjectivity weightage counts})$$

Step 3 compute sematic relevance weightage score

 If (word coocurance RSA==CSL)

 Compute the weightage score

 Semantic weightage interest score {ESc1,ESc2...}

Step 4 Return term of access doc ESc→semanticS(T)

 End

End for.

Semantic closeness between the reports is essential when it is isolated from the free substance record. Addressing the proximity and nonappearance of a thought in the twofold association may not give romanticized precision. Thought weighting through term repeat will extend the accuracy of bundled report the semantic relational score. Thought weight is made plans to use term repeat and semantic partition to categorize the documents.

3.4 Relational subset feature based clustering.

In this subjectivity of relational semantic weightage is clustered through subset clustering, they are only interested query terms retrieval is grouped by subjectivity measure from different educational documents. The data logs be categorized by group by reference which a record of query necessary searched. By computing the relative terms by fixing the cluster of latent keyword count weightage as centroid closeness to other comparison of word co-occurrence. The number of query logs accessed by entering the queries by the count word key terms in max state. The total max count have stated to the relational score to the subjectivity measure. By this similarity search cluster are group by class by reference on key terms evaluated from the educational document.

Input: Educational Docs (Edl), subjectivity lesionsSLs.

Output: Optimized clustered eds.

Step 1: compute the semantic relational weightage from SLs

For each measure group SLs from Edl

 For each query relationQrl

 Subjectivity similarity measure query Qrl→ semantic group

Step 2 Compute Number of relations by relational entity it has.

RSm→closestcentroid cluster= \sum subjectivity Relations \in SGi

Step 3: Computer Number of relative terms of access.

Step 4 Group by conceptual class.

Rls = = \sum relational concept(subjectivity Links(SGi)) \in \sum Concept(DLs)

```

Step 5: Compute group by relevance
        Max weightage measure (subjectivity score)
        Add to weight set  $Ws = \sum Ws(Edl)+concurrency$ 
        Relativity analysis query  $Qr1 \rightarrow t$  terms
        Return subjective relational subjects
for each class  $Sci$  from  $SGi \rightarrow qrl$ 
         $sci = \frac{\sum_{i=1}^{size(st)} \sum st(i).srealtionlterm ==Si \cup St(i).user ==class id}{list(\sum qrl(i).sroup ==sSi)}$ 
        Return the subjectivity class by group  $Sci$ 
    End
End
End.
    
```

Above algorithm represents the personal access based on the information of specific subjectivity accessed by learners by group by cluster evaluation of calss.

IV.RESULTS AND DISCUSSION

The research familiarizes the development educational system using latent semantic measure by cluster evaluation system with proof of web mining based learning capabilities to improve the education system. The test case results proofs are attained with semantic cluster analysis accuracy, time complexity, and false repentance with specified sensitivity and specificity measure of search logs from learners. Differential methods are evaluated for its performance in various conditions. The detail of the implementation is given below:

Table 4.1: Parameter Details of Proposed Methods.

Parameter	Value
Framework	Microsoft Visual Studio Framework
Programming Language	C#.net, java scripting
Attributes considered	Query term search attribute
Number of users	1500 learners

The above Table 1, reviews the proposed system taken the parameters to test the intent proposed methods for effective development educational management using semantic relation. The facilities are deployed in Microsoft visual studio environment, and the lookup has been performed based on the qery search approach has been implemented using c# .net and the efficiency of the method has been analyzed. The proposed plan has higher performance of test case by carrying efficient support measure, false cluster and time complexity.

Impact of educational support analysis

The educational support availability is the parameter which represents how efficient the behavioral terms is assigned on requesting to accept web education document accessed by relational query search by analyzing document content be received and how many times the particular function has been assigned.

$$\text{Behavioral accuracy} = \frac{\text{Number of subjectivity search oriented queries the Assigned}}{\text{Total number of request received}} \times 100$$

The proposed local similarity measure based relational search of term queries from learners are input through the search progress to analyze the web educational documents. The proposed has higher performance based on subjectivity analysis.

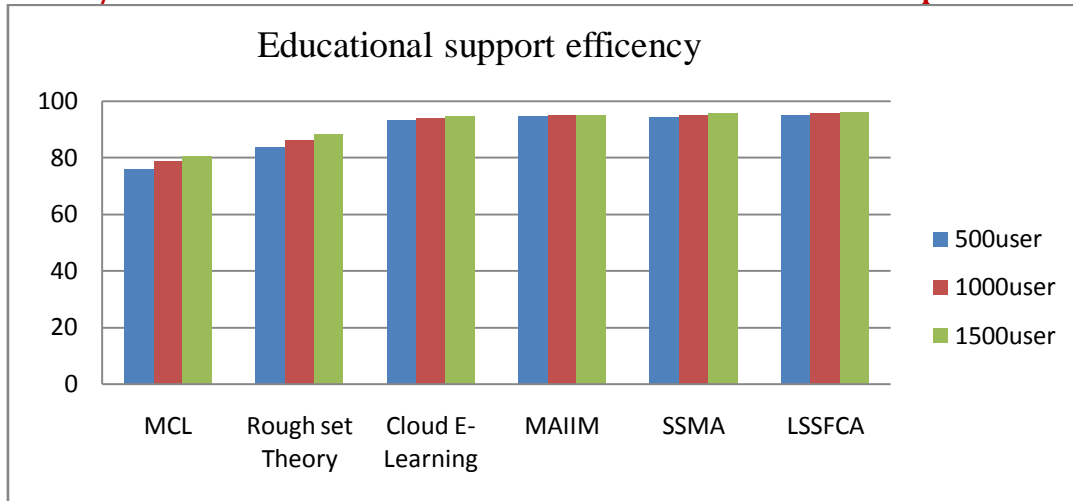


Figure 4.1: Comparison of educational support efficiency

The above figure 4.1 reviews the educational support efficiency proceeds the overall measure of semantic similarity by latent keyword term of cluster evaluation. The implementation produce higher performance compared to the existing evaluation results.

Table 4.2 Analysis of educational support efficiency

Analysis of educational support efficiency						
Methods /users	MCL	Rough set Theory	Cloud E-Learning	MAIIM	SSMA	LSSFCA
500	82.1	84.3	87.3	93.2	94.6	95.3
1000	84.3	86.2	88.2	94.1	95.2	95.6
1500	86.4	88.3	89.2	94.7	95.8	96.2

The above table 4.2 shows the overall efficiency produced by the features based cluster evaluation by deliver the education support to the students. The efficiency is calculated by average feature based cluster analysis with time of evaluation.

Impact of cluster evaluation

The proposed latent semantic similarity measure based relational search of term queries from learners. Which they want to search by relevance by accessing the subjectivity of educational resources using service acceptance eased content delivery .The results are accrued by cluster evaluation by right choice features.

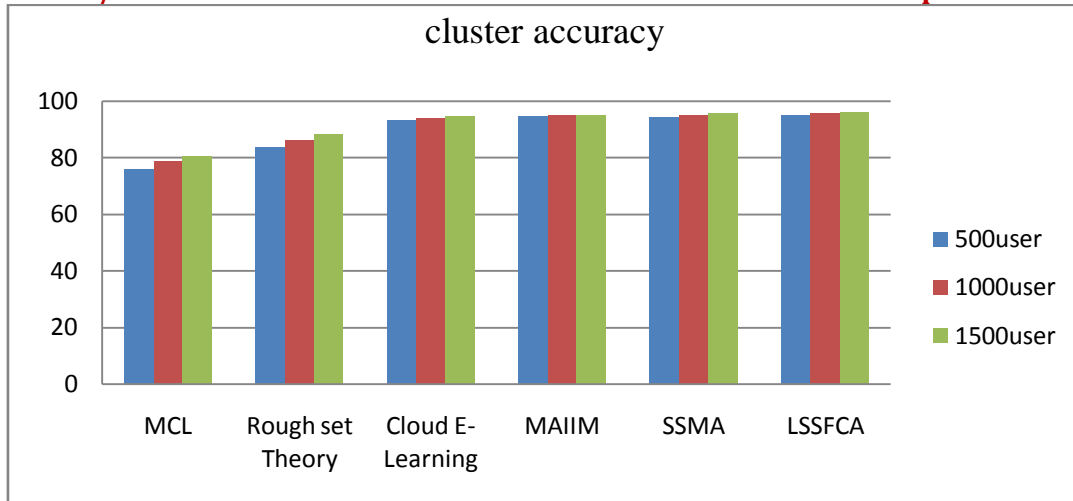


Figure 4.2: Comparison of cluster accuracy

The above figure 4.2 reviews the importance of cluster accuracy improvement has the higher performance test case result. The LSSFCA produce the cluster efficiency with the optimized features has well improved result.

Table 4.3 Analysis of cluster accuracy

Analysis of cluster accuracy						
Methods /users	MCL	Rough set Theory	Cloud E-Learning	MAIIM	SSMA	LSSFCA
500	82.1	84.3	87.3	93.2	94.6	95.3
1000	84.3	86.2	88.2	94.1	95.2	95.6
1500	86.4	88.3	89.2	94.7	95.8	96.2

The above table 4.3 shows the cluster analysis accuracy produced by LSSFCA has higher evaluation of true rate cluster form the educational content as well 95.3 % efficiency tan other methods

Impact of False analysis

The false analysis unclassified behavioral providence which in irrelevant subjectivity to the e-learners that are calculated by,

$$\text{False analysis measure FM} = \frac{\text{Number of cluster wrongly assigned}}{\text{Total number of queries}} \times 100$$

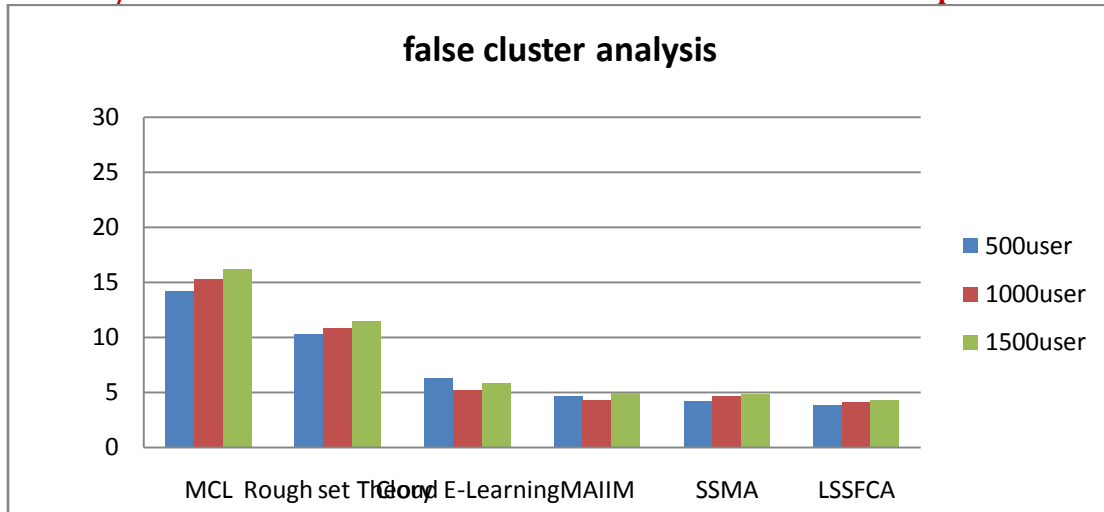


Figure 4.3: Comparison of False cluster accuracy

The above figure 4.3 reviews the evaluation of false cluster analysis achieved by best level of cluster performance. This latent semantic measures holds the least features to intent main terms to reduce the false cluster than other methods.

Table 4.4 analysis of false accuracy

Analysis of false accuracy							
Methods /users	MCL	Rough set Theory	Cloud E-Learning	MAIIM	SSMA	LSSFCA	
500	14.2	10.3	6.3	4.6	4.2	3.8	
1000	15.3	10.8	6.7	5.2	4.6	4.1	
1500	16.2	11.4	6.9	5.4	4.8	4.3	

The above table 4.4 shows the behavioral analysis accuracy produced by lower false rate because of improves cluster efficiency compared to the other dissimilar methods. The LSSFCA intends the feature to enhance the performance with lower 3.8 % false contents cluster

Impact of Time Complexity

Time complexity is analyzed to calculate the total number of time taken to execute service providence from the cloud environment to E-learners that are calculated by,

$$\text{Time complexity } T_s = \frac{\text{Number of search relational instance}}{\text{time taken for the Total number of request received}} \times 100$$

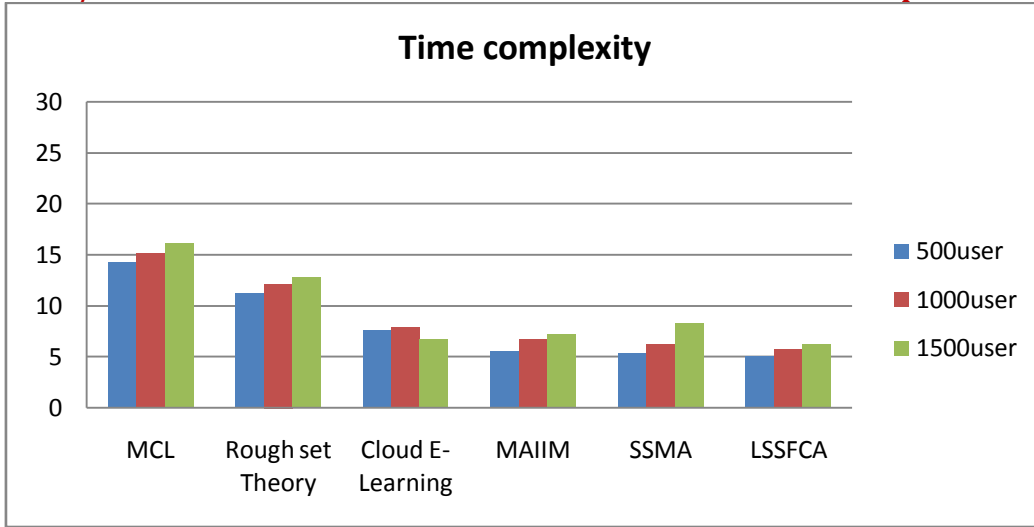


Figure 4.4: Time complexity evaluation.

The implementation test carried the above figure 4.4 shows higher efficiency with lower time competency compared to the other methods. The implementation resembles the definite of best performance than previous methods.

Table 4.5: Time competency analysis

Methods /users	Time competency evaluation (ms)					
	MCL	Rough set Theory	Cloud E-Learning	MAIIM	SSMA	LSSFCA
500	14.3	11.2	7.6	5.6	5.3	5.1
1000	15.2	12.3	7.9	6.7	6.2	5.8
1500	16.1	12.8	8.2	7.2	8.3	6.2

The above table 4.5 shows the execution of differential state of time evaluation to process the content with higher efficiency with lower complexity rate LSSFCA has 5.1 ms. The proposed system improves the time of access level in short redundant complex states that other methods.

V.CONCLUSION

To perform this, a latent semantic Similarity Measure based on feature selection algorithm is presented to improve the educational development system through relational multi-objective clustering. This method estimates the relevancy of the content in different forms by estimating relational Similarity towards text and search query towards Sentence of the document. The process improves the performance of the system in service availability up to 96.2% where the time complexity has been reduced up to 5.8 mille-seconds, and the standard measure has been achieved up to 2.6% as well. The intent measure have the higher performance improvement by choosing right features of educational development system. The proposed method has reduced the time complexity to manage efficient web service off access educational development system

REFERENCES

1. Merceron and K. Yacef, "A web-based tutoring tool with mining facilities to improve learning and teaching," in *Proceedings of the 11th International Conference on Artificial Intelligence in Education*. IOS Press, 2003.
2. D.R. Radev, H. Jing, M. Stys, and D. Tam, "Centroid-Based Summarization of Multiple Documents," *Information Processing and Management: An Int'l J.*, vol. 40, pp. 919-938, 2004.

3. L. Shen and R. Shen, "Learning content recommendation service based on simple sequencing specification," in *Proc. Int. Conf. Web-based Learning, Beijing, China, 2004*, p. 363-370
4. Y. Li, D. McLean, Z.A. Bandar, J.D. O'Shea, and K. Crockett, "Sentence Similarity Based on Semantic Nets and Corpus Statistics," *IEEE Trans. Knowledge and Data Eng.*, vol. 8, no. 8, pp. 1138-1150, Aug. 2006.
5. Pei-Chen Sun and Ray J. Tsai and Glenn Finger and Yueh-Yang Chen and Downing Yeh. What drives An empirical investigation of the critical factors influencing learner satisfaction. In *Computers and Education*, 50(4):1183 - 1202, 2008.
6. Martina Naughton, Nicola Stokes, and Joe Carthy "SentenceLevel Event Classification in Unstructured Texts" *Journal Information Retrieval archive Volume 13 Issue 2, April 2010 Pages 132-156*
7. C.Romeroans S, Ventura," Educational Data Mining: A Review of the State of the Art", *IEEE Transaction on Systems, Man, and Cybernatics*, Vol.40,No.6,2010.
8. R.M. Aliguyev, "A New Sentence Similarity Measure and Sentence Based Extractive Technique for Automatic Text Summarization," *Expert Systems with Applications*, vol. 36, pp. 7764- 7772, 2009
9. R. M. Rias and H. B. Zaman, "Understanding the role of prior knowledge in a multimedia learning application," *Australas. J. Educ. Technol.*, vol. 29, no. 4, pp. 537–548, 2013.
10. S. Ventura, A. Zafra, and P. de bra, "Applying web usage mining for personalizing hyperlinks in web-based adaptive educ. systems," *Comput. Educ.*, vol. 53, no. 3, pp. 828–840, 2009.
11. Vidhya, K. A. and Aghila, G. G. "Text Mining Process, Techniques and Tools: an Overview". *International Journal of Information Technology and Knowledge Management*. vol. 2, no. 2. pp. 613-622, 2010.
12. Gamal Ibrahim, *Budget-Aware e-Learning Systems on Cloud Computing Environments: A Genetic Approach. International Journal of Information and Educational Technology*, Vol.2, No.6, 2012
13. Vishwakarma, A.K., Narayanan, A.E." *E-learning as a Service: A New Era for Academic Cloud Approach". Recent Advances in Information Technology. IEEE, 2012.*
14. Wang, C.-C., Pai, W.-C., Yen, N.Y.: *A Shareable E-Learning Platform based on Cloud Computing. In: 3rd International Conference on Computer Research and Development (ICCRD), March 11-13, vol. 2, pp. 1–5 (2011)*
15. Andrew Skabar, and Khaled Abdalgader, "Clustering Sentence-Level Text Using a Novel Fuzzy Relational Clustering Algorithm", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 25, No. 1, 2013.
16. S. Sirkemaa, "Analysing e-Learning and Modern Learning Environments", *International Journal of Information and Education Technology*, vol. 4, No. 2, pp. 176 – 179, April 2014.
17. M. Vahdat, R. Ghio, L. Oneto, D. Anguita, M. Funk, and M. Rauterberg, "Advances in learning analytics and educational data mining," in *In: European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 2015.
18. M. Barak, R. Hussein-Farraj and Y. J. Dori, "On-campus or online: examining self-regulation and cognitive transfer skills in different learning settings." *International Journal of Educational Technology in Higher Education* 13, no. 1, pp. 35, 2016
19. MS Hasibuan, LE Nugroho,"*Learning Style Model Detection Based on Prior Knowledge in E-learning System" Second Informatics and Computing conference (ICIC),2017.*
20. Fang, C., Mu, D., Deng, Z., and Wu, Z., "Word-sentence coranking for automatic extractive text summarization", *Expert Systems with Applications*, Vol. 72, pp. 189-195, 2017.