January-February 2020
ISSN: 0193-4120 Page No. 10269 - 10278

TID Based Gap Analysis for Efficient Clinical Documentation Development using Semantic Ontology

M. Sangeetha[1], Dr. M. Senthil Kumaran[2]

[1]Research Scholar, SCSVMV University, Kanchipuram
[1]Assistant Professor, SRMIST, Kattankulathur, Kancheepuram
[2]Department of Computer Science and Engineering, SCSVMV University, Kanchipuram

Article Info
Volume 82
Page Number: 10269 - 10278
Publication Issue:
January-February 2020

Article History
Article Received*: 18 May 2019*
Revised*: 14 July 2019*
Accepted*: 22 December 2019*
Publication*: 19 February 2020*

Abstract

The problem of clinical document gap analysis has been well studied. There are numerous techniques identified which uses only the Meta data of the document. However, the methods produces higher false rate in gap analysis. To overcome the deficiency, this paper presents a Topical Informative Depth based gap analysis in clinical documentation. Instead of looking for completeness of all the fields of a topic, the TID measure represent the depth of information about any context can be used. The method reads the patient chart and records first. In the second stage, according to the category of the disease, the list of features, symptoms, diagnosis, test results are obtained from Meta data. The record has been verified for the presence of all the features considered. Based on the features identified, the method estimates the semantic coverage score, informatics coverage score. Using these two measures, a Topical Informative Depth measure for the document is estimated. According to the TID value, the completeness of the document has been verified. The method produces higher accuracy in clinical document gap analysis.

Index Terms: *Gap Analysis, Disease Prediction, Decisive Support, TID, Semantic Ontology*

1. Introduction

The modern society has higher implication of various diseases. As, there are number of new diseases has been identified every year, it is necessary to predict the possible diseases based on some symptoms. However, the presence of any disease can be identified based on set of symptoms, the most diseases has same set of symptoms. This increases the challenge for the medical practitioner to make decision on the type of disease. To support the medical practitioners, different decisive support systems has been designed. The decision making in medical industry is the process of identifying the closest disease for the input pattern which is performed based on the previous history.

So, the trace of patients is more important and it supports the growth of decisive support systems. Whatever the solution, the record and the proof is more important in making decisions. The clinical documentation is the one which makes the patient data as a proof for the decisive making systems.

The clinical documentation is the process of proof reading the patient record in form of document. The patient medical information has been generated as document. Whatever the information presents in the medical data set, regarding the patient has been converted into electronic document. The medical document would

January-February 2020
ISSN: 0193-4120 Page No. 10269 - 10278

contain information's like personal, diagnosis, treatment, treatment results and so on. Such information's has been converted into clinical document which has been used by various industries like insurance, medical research, decisive support and so on. The insurance industry would use the clinical document to make decisions on releasing funds in case of any death or accident. So, the clinical documents become proof at different situations.

The clinical documentation has higher importance in many situations like providing exact treatment and monitoring the patient health. Also, it has higher importance in claiming medical reimbursement from insurance companies. However, the medical organizations maintain various information about the patient, maintaining document in proper form is more essential. Because, it has been used for several purpose like medical analysis, health departments of government, research sectors. The hospitals have the responsibility of maintaining healthy records about the patients. For any clinical document, there are many factors to be considered. The main factor is the completeness of the document. The document would have huge data but the completeness is about the presence of necessary fields. Providing lot of textual contents is not necessary but displaying the necessary information is highly essential. This research is about generating clinical document in complete form.

There is number of gap analysis available which uses contextual information of the documents. For example, if the document is describing the information about a diseased, then it has to discuss all the necessary information. But the previous method does not monitor the presence of all the fields of the disease. For example, if the document is about a diabetic, then it has to display the details of various diagnosis like "blood glucose; body mass index; pre meal glucose; post meal glucose; insulin; creatinine" and so on. Also, the document has to display other features like the lifestyle and food habits of the patient. The previous method does not consider and verify the presence of all the necessary fields of information in publishing them.

Table 1: Sample data set

Patient ID	Temp	Body Pain	Stomach Pain	Vomiting	Disease	Treatment	Result
A	101	Yes	No	No	Fever	Paracetamol	Yes
B	100	Yes	Yes	Yes	Typhoid	Cefixime Of laxacin	Yes
C	101	Yes	No	No	Fever	Paracetamol	Yes
D	101	yes	yes	yes	Typhoid	CefiximeOflaxacin	yes
E	100	yes	yes	yes	Typhoid	CefiximeOflaxacin	No

The Table 1 shows the details of diagnosis and treatment provided and maintained by any medical organization. The details available in Table 1 can be used to generate clinical documentation and the same can be used to perform gap analysis.

<Ontology>
<patient>
<ID>
<Name>
<Age>
<Sex>
<Diagnosis>
<Blood Test>
<Widal>
<HbA1C>
<Blood Sugar>
<Treatment>
<Medicines>
<result>
</Ontology>

The above presented ontology has been maintained and would be used to perform gap analysis. Any clinical document could be verified for the presence of all the features and it is not necessary that the document should contain all the features of the ontology, but it should cover the mandate features. Semantic ontology has been applied for several issues. It can be used for the problem of gap analysis. The semantic ontology would contain number of features belongs to any class of disease and

Would contain many properties related to person, disease, treatment and so on. It can be used to perform gap analysis.

Towards the motivation, a Topical Informative Depth measure based approach is presented in this paper. The method considers both textural features and information related to disease. Using the features the method estimates depth measure to perform gap analysis. By verifying the depth of information the document covers and depth of features it covers, the problem of gap

January-February 2020
ISSN: 0193-4120 Page No. 10269 - 10278

analysis can be performed efficiently. The detailed approach is discussed in the next section.

2. Related Works

There are number of approaches discussed towards gap analysis. This section discusses set of methods related to the problem.

In [1], the author claims that, it is important to analyses the gap in the documents of medical drugs, medical device, medical procedure, or disease/therapeutic area. A gap analysis assesses how a drug, for example, is represented in the medical literature and at scientific congresses and how current coverage of the drug relates to a company's internal goals or competitor products. By analyzing the gap in the documents of drug, it can be verified that the document covers all the necessary information. The method considered factors like identifying the area(s) of focus/rationale, determining a meaningful timeline for analysis, identifying the scope of research, determining search parameters, selecting a format for gap analysis output, conducting the search, and organizing and prioritizing findings to identify trends regarding the topic question. Systematically following these steps will expose knowledge gaps about a drug product, medical device, medical procedure, or therapeutic area in the current literature that can then be addressed with targeted publications as part of a medical publication plan.

In [2], the author presents a tool which has been used to identify the gap available in the documents towards patient's expectation. The tool has been used to perform analysis on various perspective and focused towards the understanding of nurses. Similarly in [3], the author developed a gap analysis algorithm which uses health records, policies. The method identifies the gap present in the document and helps to perform diabetic management.

In [4], an evidence based gap analysis system is presented for the decision support. The method uses case specific information, clinical details and information obtained from medical experts. The system has been used to mine useful information from huge data set and has been used to identify the gap present.

In Robustness, fidelity and prediction-looseness of models [5], various mathematical and numerical models are used to identify gap in the result of the prediction models. The method identifies various looseness in the prediction and fidelity of the data and their robustness to varying data. It has been performed by measuring mean square error, fidelity and so on. Similarly in [6], a decision making approach is presented for the disease prediction of animals which considered the exotic infections in them. The disease has been occurring to them at the transportation and they have not been tested due to the cost. The method performs detailed review on various methods available for the prediction of the disease and performs a gap analysis.

In [7], the author present an approach for the gap analysis of documents related to health care systems. Various approaches of gap analysis are considered and the data has been extracted from different clinical data. The gap analysis is performed by considering different health care data.

In [8], the author present a pattern based mining algorithm for health records using clinical analysis. Different patterns of the document have and health records are maintained. Based on the pattern, the presence of all features of the records in the document has been validated. Based on the pattern matching the gap analysis is performed.

In [9], a comprehensive study is performed on the knowledge development of medical industry. The method applies various data mining techniques to identify clinical data from patient data set. The method has been validated using various cancer data set.

In [10], the author presents a gap analysis technique which uses text mining techniques and statistical results. The method uses different mining techniques and visualization tools.

In [11], the author discusses the implication of text mining approaches towards converting electronic health records to documents. The method uses various text mining techniques and natural language processing techniques has been used to perform gap analysis. The application of NLP techniques has improved the performance of gap analysis.

In [12], which extract the semantic terms from the corpus and measure the semantic similarity with different categorical disease. If any of the disease class is identified more closure but misses with set of features, it has been identified as gap.

In [13], for the prediction of heart disease the method used lexical rules which use clinical notes. The method uses natural language processing and clinical data for the prediction. The prediction is performed by generating lexical rules from the clinical data. The method has produced efficient results on gap analysis. Similarly in [16], the document classification is performed using NLP techniques. The UIMA is focused towards classification using data driven approach and focused towards classification.

In [14], the author present a heart sound classification algorithm which receives the heart sounds

January-February 2020
ISSN: 0193-4120 Page No. 10269 - 10278

through the motifs. The method classifies the heart sounds into different classes like murmur, normal and extracts systole using SAX approach. The method performs preprocessing to convert the sound into SAX operands. The classification is performed by computing the term frequency and inverse document frequency measures for each document.

In [15], a personalized prediction model has been presented for the support of clinical decision making which uses coronary syndrome details. The decision support has been considered for the Acute Coronary Syndrome (ACS). The method has adapted different decision making approach to be combined to produce the clinical decision making. The method has used state and time based data to perform prediction.

All the above discussed methods suffers to produce higher accuracy in gap analysis.

TID Based Clinical Document Gap Analysis

The proposed method reads the input clinical document set. From the clinical document, the method extracts textual features. Using the text features extracted, the method uses natural language processing technique to extract key terms. Based on the features extracted, the method estimate semantic coverage scores on different class of semantic ontology. Similarly, the method estimates the informatics coverage score. Using these two measures, the method estimates the TID value to perform gap analysis. The detailed approach is presented in this section.

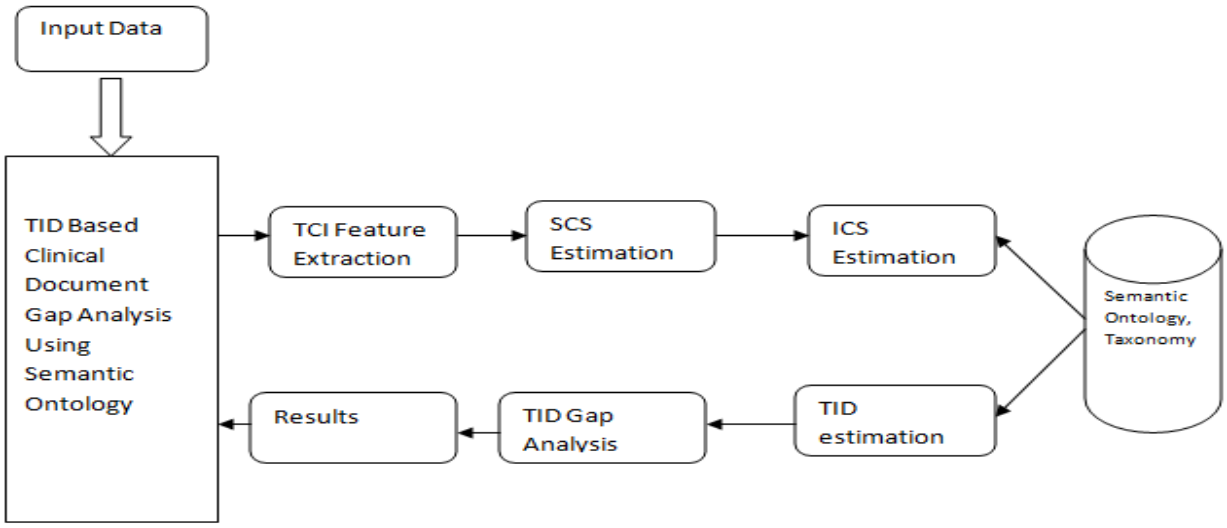

Figure 1: Architecture of TID Based Gap Analysis

The architecture of proposed TID based approach is presented in Figure 1. The functional components of the proposed method have been discussed in detail in this section.

TID Feature Extraction

This is a preprocessing stage of the gap analysis which reads the input document set given. From each document D from the set Ds, the methods extract the content features like text. The features being extracted has been split into term set. It is followed by a stop word removal process. From the stop word removed content, the list of key words has been identified. For each term from the term set, the method computes the subject value measure (SVM) and Informative value measure (IVM). The SVM represent the membership of the document under the class of any disease where IVM represent the value of the text towards any parameter of diagnosis or other values. It has been measured based on the taxonomy available and the ontology maintained.

Algorithm

Input: Document Set Ds, Domain Ontology O
Output: Feature Set Fss, FIs.
Start
Read document set Ds.
Read
For each document Di

January-February 2020
ISSN: 0193-4120 Page No. 10269 - 10278

Extract text feature Tf = $\int_{i=1}^{size\,(Di)} \sum Text \in Di$

For each term Ti

If $\int_{i=1}^{size\,(swl)} Ti \in swl(i)$ then

Tf = $\sum(Terms \in Tf \cap Ti$

End

End

For each term Ti

Compute subject value measure SVM.

$$SVM = \frac{\sum_{i=1}^{size\,(O)} O(i) \in Ti \,/\, Size\,(O)}{\sum Terms\,(Tf)}$$

Compute Informative Value Measure (IVM)

$$IVM = \frac{\sum_{i=1}^{size\,(Tf)} Tf(i) \nexists \forall (O) \,/\, size\,(Tf)}{Number\ \ of\ \ classes}$$

If SVM > IVM then

Fss = $\sum Terms(Fss) \cup Ti$

Else

FIs = $\sum Terms(FIs) \cup Ti$

End

End

Stop

The above discussed algorithm extracts the features from the document and identifies the subject and value features. Extracted features have been added to the feature set which has been used to perform gap analysis.

SCS Estimation

The subject coverage score represent how good the document covers the subject. For example, consider the document is related to the disease 'Cardiac Malfunction'. In this case, the document has to discuss more about cardiac issues. Also it should cover many features like pressure, age, sugar, cholesterol, regular ECG, number of attacks faced and so on. If the document does not speak about more information related to the disease identified, then it can be mentioned as irrelevant document. In order to fall within the specific category, the document has to discuss more about the subject information. It has been measured based on the SCS value. It has been measured using the terms available in the document.

Algorithm

Input: Feature set Fss, Ontology O, Taxonomy T
Output: SCS.
Start

Read feature set Fss.

For each Disease class c

Identify list of all class features.

CFL = $\int_{i=1}^{size\,(O(c))} \sum Features(O(c))$

Identify list of Topical features.

TFL = $\int_{i=1}^{size\,(T(c))} \sum Topical\ Features(T(c))$

Compute SCS = $\frac{\sum_{i=1}^{size\,(CFL)} CFL(i) \in Fss}{size\,(CFL)} \times \frac{\sum_{i=1}^{size\,(TFL)} TFL(i) \in Fss}{size\,(TFL)}$

End

Stop

The above discussed algorithm estimates the subject coverage score for different category identified. Estimated SCS value has been used perform gap analysis.

ICS Estimation

The informative coverage score is the measure which represents the amount of information related to the subject features available in the document. Even though the document speak about the topic, it is necessary to cover set of features and their values. For example, when the document speak about Typhoid, it is necessary to present the values of diagnosis results and their values in the document. Also it should cover the type of treatment, medicines given and the result of treatment. All these information values has to be present in the document. It has been measured based on the feature set extracted and the semantic ontology available.

Algorithm

Input: Feature set Fis, Ontology O
Output: ICS.
Start

Read feature set FIs.

For each Disease class c

Identify list of all value features.

VFL= $\int_{i=1}^{size\,(O(c))} \sum Features(O(c)(Values))$

Identify list of value features with value.

VFLV= $\int_{i=1}^{size\,(O(c))} \sum Features(O(c)(Values)! = Null)$

January-February 2020
ISSN: 0193-4120 Page No. 10269 - 10278

Compute ICS = $\frac{VFL}{size\,(FIS)} \times \frac{VFLV}{size\,(VFL)}$

End

Stop

The above discussed algorithm computes the informative coverage score for different class of diseases. Estimated ICS value has been used to perform gap analysis.

TID Gap Analysis

The topical informative depth based clinical document gap analysis algorithm reads the input document set. For each document, the method extracts the topical and informative features. Using the features extracted, the method estimates subject coverage score (SCS) and informative coverage score (ICS) values. Using the measures estimated, the method computes Topical Informative Depth (TID) measure. Estimated TID measure has been used to classify the document as complete or incomplete. If the TID value of any document is greater than threshold it is considered as complete otherwise it has been considered as incomplete.

Algorithm:

Input: Document Set Ds

Output: Null

Start

Read Document set Ds.

For each document Di

[Fss,FIs] = TIFeatureExtraction (Di)

Compute SCS = SCS-Estimation (Fss, Ontology)

Compute ICS = ICS-Estimation (FIS, Ontology)

Compute TID = SCS×ICS

If TID>Threshold

Complete

Else

Incomplete.

End

End

Stop

The above discussed algorithm computes the topical informative depth measure towards each class of disease. According to the value of TID measure, the method decides the fitness of the document and classifies them.

3. Results and Discussion

The proposed TID based clinical document gap analysis algorithm has been implemented using advanced java. The method has been implemented and evaluated for its performance in various parameters of gap analysis. The method has produced higher efficiency in gap analysis and produces less classification ratio. The details of data set being used to evaluate the performance of the algorithm are presented below:

Table 2: Details of Data set

Parameter	Value
Data Sets	Indian Liver Patient Data set (ILPD), AstraZeneca
Number of Data Points	583,606
Dimension of Tuples	10, 9

The details of data set being used for the evaluation of the proposed algorithm have been presented in Table 2. The method used ILPD data set, which present the records of Indian liver patients which has 10 numbers of attributes and in total there are 583 number of records available. The data set contains the information like age, gender, total Bilirubin, direct Bilirubin, total proteins, albumin, A/G ratio, SGPT, SGOT and Alkphos. Similarly, Astrazeneca data set has been used with 606 numbers of data points which covers 9 parameters. It covers the features of medicines given for the treatment of liversirosiz. These two data sets has been merged to produce large number of data points with more features. Cooked up data set has been used to evaluate the performance of the algorithm proposed.

According to the above details, the snapshot of data set has been presented below:

Table 3: sample data set considered

Age	Sex	TB	DB	Alkphos	Sgpt	Sgot	TP	ALB	A/G
65	Female	0.7	0.1	187	16	18	6.8	3.3	0.9
62	Male	10.9	5.5	699	64	100	7.5	3.2	0.74
62	Male	7.3	4.1	490	60	68	7	3.3	0.89
58	Male	1	0.4	182	14	20	6.8	3.4	1
72	Male	3.9	2	195	27	59	7.3	2.4	0.4
46	Male	1.8	0.7	208	19	14	7.6	4.4	1.3

January-February 2020
ISSN: 0193-4120 Page No. 10269 - 10278

26	Female	0.9	0.2	154	16	12	7	3.5	1
29	Female	0.9	0.3	202	14	11	6.7	3.6	1.1
17	Male	0.9	0.3	202	22	19	7.4	4.1	1.2
55	Male	0.7	0.2	290	53	58	6.8	3.4	1
57	Male	0.6	0.1	210	51	59	5.9	2.7	0.8

The sample data records of the data set have been presented in Table 3. According to the data set, the medical documentation has been generated using publication schemes. The generated clinical document has been validated for its accuracy using the proposed algorithm.

Similarly, the liver cirrhosis disease related documents have been considered for gap analysis. The documents are measured for different measures according to the proposed TID scheme. Consider the document text as follows:

Table 4: Sample Clinical Document

Patient Details:
Name : x Age: 32
Diagnosis:
The patient x has been Tested for Liver cirrhosis. He have been undergone for different tests and diagnosis at scan, blood test and so on. Each report has been considered and recommended set of treatments. The details of treatment has been presented in the next section.
Blood Test:
HbA1c, Cholestrol, Albumin
Treatment:
X,Y,Z
Result:

The sample clinical document considered for the evaluation is presented above. The document contains various features and details. According to the document text, various measures are estimated to perform gap analysis.

The text feature extracted from the Table 4 content can be performed with TI Feature Extraction. The feature extraction algorithm would identify list of key terms by removing the stop words. The result of feature extraction can be mentioned as follows:

Table 5: Stop word removed content

Stop word removed content
Patient Details
Name x Age 32
Diagnosis
The patient x Tested Liver cirrhosis undergone different tests diagnosis scan blood test report considered recommended treatments treatment Blood Test
HbA1c Cholestrol Albumin
Treatment
X Y Z
Result
Liver Cirrhosis

The text features after removal of stop words is presented in the above Table 5. The text features are read and the presence of stop words has been removed and identified for list of key terms. Identified terms have been used to perform different measures.

Table 6: Subject and Informative Features

Subject Features	Informative Features
Liver Cirrhosis	Scan blood test HbA1c, Cholesterol, Albumin Treatment X,Y,Z Result

The subject and informative features present in the document has been extracted and presented in Table 6. Using the features extracted, the method estimates the SCS and ICS measures.

Gap Detection Accuracy

The proposed method has been measured for the accuracy in gap detection over the submitted document. The measured result has been compared with the results of other methods. The gap detection accuracy has been measured as follows:

January-February 2020
ISSN: 0193-4120 Page No. 10269 - 10278

Gap Detection Accuracy = $\frac{TPC+TFC}{Total\ Number\ of\ Documents}$

Table 7: Performance on gap detection accuracy

Gap Detection Accuracy			
Method	30 Features	50 Features	100 Features
Text Mining	75	71	69
Pattern Based	81	75	74
UIMA	86	79	76
TID	98	97	95

The accuracy on gap detection has been measured for the proposed method and compared with the result of other methods. The proposed TID based approach has produced higher gap detection accuracy at varying feature levels and feature lengths.

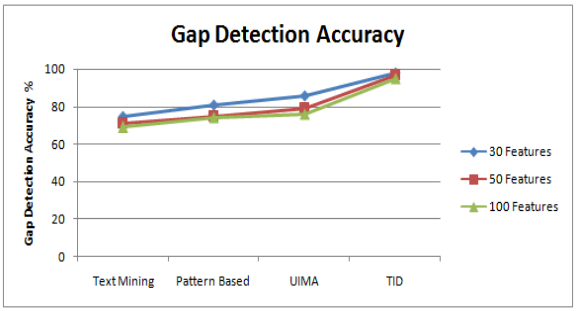

Figure 2: Comparison on Gap Detection Accuracy

The performance on gap detection accuracy has been measured for different algorithms. The proposed TID based gap analysis algorithm has produced higher performance than other methods.

False Classification Ratio

The proposed method has been measured for the false classification ratio in gap detection over the submitted document. The measured result has been compared with the results of other methods. The false classification ratio has been measured as follows:

False Classification Ratio = $\frac{TNC+TPC}{Total\ Number\ of\ Documents}$

Table 8: Performance on False Classification Ratio

Gap Detection Accuracy			
Method	30 Features	50 Features	100 Features
Text Mining	25	29	31
Pattern Based	19	25	26
UIMA	14	21	24
TID	2	3	5

The ratio of false classification of document has been measured for different methods at varying number of features. The proposed TID algorithm has produced less false ratio up to 5 % which is less than other algorithms.

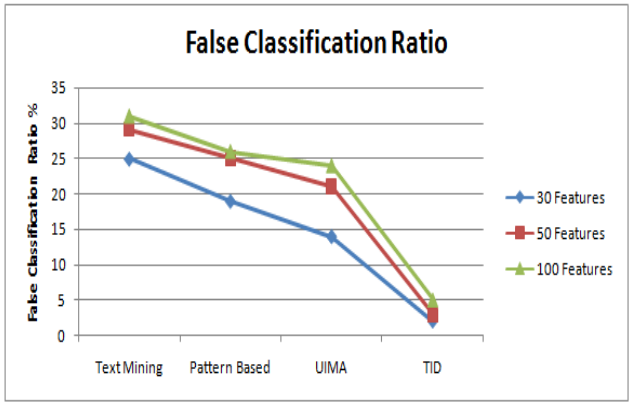

Figure 2: Comparison on False Classification Ratio

The performance on false classification ratio has been measured for different algorithms. The proposed TID based gap analysis algorithm has produced less false ratio than other methods.

Time Complexity

The time complexity is the factor which represents the rapidness of the algorithm in classifying the document. As the document size is higher, the time complexity plays vital role in the performance measurement of any algorithm. It has been measured as follows.

Time Complexity = Total Time taken for the classification of document.

Table 9: Performance on Time Complexity

Time Complexity			
Method	30 Features	50 Features	100 Features
Text Mining	45	57	78
Pattern Based	39	53	72
UIMA	34	42	68
TID	23	32	43

Time complexity in gap analysis at varying feature size has been measured for different algorithms and presented in Table 9. The proposed TID algorithm has reduced the time complexity than any other algorithm.

January-February 2020
ISSN: 0193-4120 Page No. 10269 - 10278

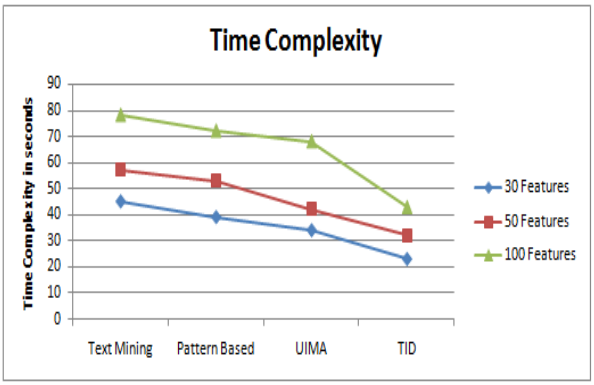

Figure 3: Comparison on Time Complexity

The performance on time complexity has been measured for different algorithms. The proposed TID based gap analysis algorithm has produced less time complexity than other methods.

4. Conclusion

In this paper, an efficient clinical document gap analysis algorithm is presented. The method reads the input document set and for each document, the text features and informative features are extracted. Using the features extracted, the method estimates subject coverage score and informatics coverage score. Using the measures estimated, the method computes topical informatics depth measure. Based on the value of TID measure, the method identifies the gap present in the document. The proposed algorithm improves the performance of gap detection accuracy and reduces the false classification ratio.

References

[1] Finucane, Stephanie, Conducting a Gap Analysis for a Medical Publication Plan, AMWA Journal: American Medical Writers Association Journal. Fall, 29, 3,104-110,(2014).

[2] Rhian Silvestro, "Applying gap analysis in the health service to inform the service improvement agenda", International Journal of Quality & Reliability Management, 22, 3,215-233, (2005).

[3] Sherita Hill Golden, and MD, MHS; Daniel Hager, MHA; Lois J. Gould, MS, PMP; Nestoras Mathioudakis, MD, MHS; Peter J. Pronovost, MD, A Gap Analysis Needs Assessment Tool to Drive a Care Delivery and Research Agenda for Integration of Care and Sharing of Best Practices Across a Health System, The Joint Commission Journal on Quality and Patient Safety, 43, 18–28,(2014).

[4] Kari Sentz, François M. Hemez, Information Gap Analysis for Decision Support Systems in Evidence-Based Medicine, International Workshop on Machine Learning and Data Mining in Pattern Recognition MLDM,543-554,(2013)

[5] Ben-Haim, Y., Hemez, F.M.: Robustness, fidelity and prediction-looseness of models. Proceedings of the Royal Society A. 468, 2137, 227–244 (2012).

[6] Troffaes, M.C.M., Gosling, J.P.: Robust Detection of Exotic Infectious Diseases in Animal Herds: A Comparative Study of Three Decision Methodologies under Severe Uncertainty. IJAR (2013)

[7] Anke Rohwer and Anel Schoonees, Taryn Young, Methods used and lessons learnt in conducting document reviews of medical and allied health curricula – a key step in curriculum evaluation, BMC Medical Education, 14, 236 (2014).

[8] Oleg Metskera, and Ekaterina Bolgovaa, Alexey Yakovlevab, Pattern-based Mining in Electronic Health Records for Complex Clinical Process Analysis, Elsevier, Procedia Computer Science, 119, 2017, 197-206 (2017).

[9] V. Rakocevic, and G. Djukic, T. Filipovic, N. Milutinovic Computational medicine in data mining and modeling, Springer, New York (2013)

[10] B. Miner, and G., Delen, D., Elder, J., Fast, A., Hill, T., &Nisbet, Practical text mining and statistical analysis for non-structured text data applications. Amsterdam: The Netherlands: Academic Press, (2012).

[11] E. H. Hansen, and B. Bruun, L. Firm, P. Warrer, E. H. Hansen, and L. Juhl-jensen, "Using text-mining techniques in electronic patient records to identify ADRs from medicine use using text-mining techniques in electronic patient records to identify ADRs from medicine use," no. September (2011).

[12] Z.V. Mitrofanova "Automatic analysis of terminology in the Russian corpus on corpus linguistics" /Computer Linguist. Intellect. Technol. Proc. Annu. Int. Conf. "Dialogue", 321-328,(2009)

[13] G. Karystianis, and A. Dehghan, A. Kovacevic, J. A. Keane, and G. Nenadic, "Using local lexicalized rules to identify heart disease risk factors in clinical notes," 58, (2015).

[14] E. F. Gomes, and A. M. Jorge, L. Tec, P. J. Azevedo, and H. Tec, "Classifying Heart Sounds

January-February 2020
ISSN: 0193-4120 Page No. 10269 - 10278

using SAX Motifs, Random Forests and Text Mining techniques," 334–337, (2014).

[15] A. V Krikunov, and E. V Bolgova, E. Krotov, T.M. Abuhay, A.N. Yakovlev, S. V Kovalchuk "Complex data-driven predictive modeling in personalized clinical decision support for Acute Coronary Syndrome episodes" Procedia Computer Science, 80, 2016, pp. 518-529, (2016).

[16] Y. Heights "UIMA : an architectural approach to unstructured information processing in the corporate research environment", 327-348, (2017).

[17] Yu-Cheng Lee, and Yu-Che Wang, Chih-Hung Chien, Chia-Huei Wu, Shu-Chiung Lu, Sang-Bing Tsai, Weiwei Dong, Applying revised gap analysis model in measuring hotel service quality Springerplus. 5,1 (2016).

[18] Bahae Samhan; K.D. Joshi, Resistance of Healthcare Information Technologies; Literature Review, Analysis, and Gaps, Hawaii International Conference on System Sciences, (2015).