International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-5, March 2019

Min Feature Replacement Algorithm Based Multiple Imputation Based Gap Analysis for Clinical Data

M Sangeetha, M Senthil Kumaran

Abstract*: *The impact of technology has expanded its wings in many fields and medical domain has great impact from technology development. The technology development has been enforced in medical organization to convert the patient details from the charts to electronic form of health records. The patient records have numerous information and not all of them are important and they have been visited for specific purposes like medical reimbursement from insurance corporations. There are departments in medical organization which consists of doctors, specialists and nurses. They always look for the precise information regarding the patient about their diagnosis results and treatments. They focus on generating documentation in efficient way which provides necessary information in accurate manner. The problem of clinical document improvement has been well studied. There are number of issues has been identified from the way the clinical documents has been published. This work, present a min feature replacement based multiple imputation technique for efficient gap analysis. The method reads the input data set and identifies the list of features and estimates the minimum feature values. Second for each record, the method identifies the missing values, and replaces them with min feature value identified. The replaced imputed data set has been used to perform disease prediction. The method produces higher performance in multiple imputations and improves the performance of prediction.

Index Terms*: *Multiple Imputations, Missing Values, HCS, CDI, Gap Analysis, Publication.

I. INTRODUCTION

The modern human society has been affected by number of diseases which are known/unknown and identified every day. Diagnosing the prone disease becomes a highly challenging task for the medical practitioner. So, they need the support of medical documents to get the conclusion and the knowledge about the disease and the treatment possibilities available. The medical documents are published by various organizations towards the health control. The HCS (Health Care Systems) are the units of medical organizations which supports the controlling of diseases. They really work over the diseases based on the information available in the documents. The most medical practitioner looks for the medical documents to get the knowledge about any issue. So it is necessary for the document to provide complete information on the specific issue.

On the other side, the EHR (Electronic Health Record) maintained by any organization would have missing values. Such missing values would affect the performance of prediction performed or outcome of the prediction process. So multiple imputations are performed to fill the missing values. It has been performed by generating duplicate rows of data points with cookeup values. The value being replaced has been performed in several ways. Identifying the missing values has been named as gap analysis. The gap analysis is the process of identifying the possible gap available in the document. It is not necessary that the medical document should provide entire information on the disease considered. There are documents which specify only outlines and may not discuss entire information. In the medical practitioner point of view, the document should discuss the entire details. For example, consider the case of “breast Cancer”, if the document is discussing about the disease, then it has to discuss various factors with information and evidences. In this case, the document has to discuss about the symptoms, risk factors, treatment options, diagnosis options, history of patients, evidences, drug details, remarks and the treatment results. All these information should be discussed in the document to become a complete one. When the document misses any of the feature above mentioned has becomes a incomplete one.

At the time of publication, the documents have to be verified for their completeness. The publication is the process of uploading the data to the view of external world which needs to have all the metrics above discussed. In [17], the application of revised gap analysis towards the service measurement of hotels is discussed. The method is proposed to monitor the quality of service in hotel industries.

In [18], the author performed a detailed review on gap analysis. The author performs analysis of different approaches of gap analysis towards HIT resistance. The health care systems have been considered for the development and medical documents have been considered for gap analysis.

II. LITERATURE SURVEY

There are number of methods have been discussed for the problem of gap analysis in medical documents. Some of the methods have been discussed here in this section.

Manuscript published on 30 March 2019.
***Correspondence Author(s)
M Sangeetha, Department of of Information Technology, SRM Institute of Science and Technology, Chennai, India.
M Senthil Kumaran, Department of Computer Science and Engineering, SCSVMV, Kanchipuram, India.

2.1 Text Mining Techniques on Gap Analysis

The text mining techniques can be well adapted to the problem of gap analysis. The text mining techniques always search for the terms in the document given. They measure the similarity of the document towards any class. The similarity measure is estimated based on the occurrence of the terms of the category considered. In gap analysis, if the terms and key words of the particular category have been well known, then they can be searched through text analysis methods. This would estimate the fitness of the document which specifies the presence of gap in the context.

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

2.2 Pattern Mining Techniques in Gap Analysis

Any health care data have number of information. So for any health care document, it is necessary to contain all the features identified. Such identified features can be framed as a pattern. The document feature can be extracted and verified with the pattern. It would help to identify the gap present in the document.

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.

2.3 Prioritized Approach on Gap Analysis

The health care documents would discuss about many features of any disease and the supporting documents would discuss various information. Also, according to the patients point of view, they would have different expectation on how the medical document should be. So they have certain priority in representing the medical document. Such methods are discussed in this section.

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 perspectives and focused towards the understanding of nurses. Similarly in [3], the author developed an gap analysis algorithm which uses health records, policies. The method identifies the gap present in the document and helps to perform diabetic management.

2.4 Decisive Support Systems on Gap Analysis

The decisive support systems are the main component for the medical practitioner in making decisions. In many situations, the medical practitioner would suffer to make any decisions. At this stage, they approach the decisive support systems which would help them. The DSS are trained using the health care documents. By extracting the features from the health care documents, various information of the medical record can be extracted and trained towards number of disease classes. So, in order to index into a category, it is necessary for the document should contain all the feature values required. This type of systems performs gap analysis by verifying the presence of all the features of the class identified.

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 [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.

2.5 Numerical Models on Gap Analysis

Various numerical models have been used for the problem of gap analysis in medical documents. Such methods are discussed in this section. 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.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

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 is considered and the data has been extracted from different clinical data. The gap analysis is performed by considering different health care data.

2.6Linguistic Approaches in Gap Analysis

The linguistic approaches are performs the gap analysis based on the semantic meanings. The natural language processing techniques has been used to identify the related terms of any category. By maintaining the semantic ontology of various disease classes, the presence of the terms related to semantic classes can be identified using the NLP techniques and linguistic approaches. This would improve 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 [19]. 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.

III. MIN FEATURE REPLACEMENT BASED MULTIPLE IMPUTATION

The proposed multiple imputation algorithm reads the input data set. From the input data set, the method identifies the list of features. Based on the list of features identified the method reads each input data point, and verifies for the presence of all features. In the second stage, the method identifies for the presence of values in each data feature identified. If there is a missing value for any feature, then it has been considered as missing values. The missing values have been replaced by min feature value generated. The detailed approach is discussed in this section.

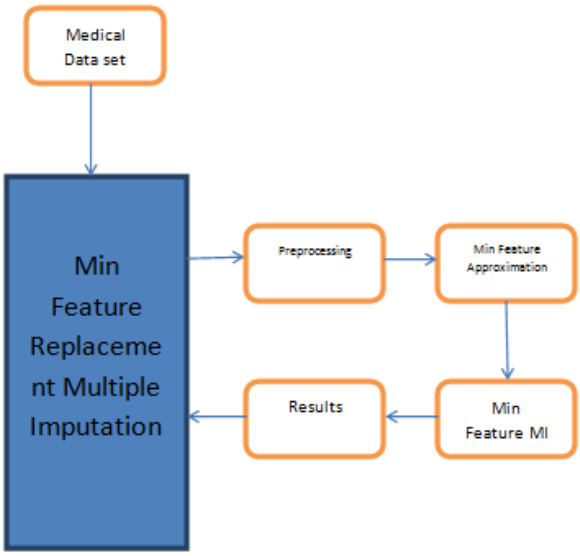

Figure 1: Architecture of Min Feature MI

The Figure 1 shows the architecture of proposed min feature multiple imputation system and shows the functional components in detail.

Preprocessing:

In this stage, the method reads the input data set. For each data point Di, the method identifies list of dimensions available. The method identifies the distinct dimensions from the data set. Identified dimension list are used for min feature approximation.

Min Feature Approximation:

In this stage, the method reads the data set and for each feature of the data set, the method read the feature values. Based on the feature values, the method estimates the minimum and average values. Based on these values, the method estimates the min feature value Mf. Computed Mf value has been used for multiple imputation.

Min Feature MI:

In this stage, the method reads each data point. For each data point, for each dimension or feature, if there is no value present then it has been replaced by the Mf value of the dimension estimated earlier. The imputed data set has been used for further processes.

Algorithm:

Input:Data set Ds
Read data set Ds.
Identify list of dimensions Ldim.

$$Ldim = \int_{i=1}^{size(Ds)} \sum Dimension(Ds(i) \nexists Ldim)$$

For each dimension di
Compute minimum feature value Mfv. $Mfv = \int_{i=1}^{size(LDim)} Min(\sum Ds(i)$

Compute mean feature value Mefv.

$$Mefv = \frac{\sum_{i=1}^{size(Ds)} Ds(i)(di)}{size(ds)}$$

Compute imputation min feature Imf.

$$Imf = Mfv + \frac{Mefv}{Mfv}$$

End

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

For each data point Di
For each dimension Dim
If(Di(Dim)==Null)
Perform multiple imputation MI.
Replace Di(Dim) with Imf(Dim).
End
End
Stop.

The above discussed algorithm performs multiple imputations on the data set given according to the missing values based on min feature approximation techniques.

IV. RESULTS AND DISCUSSIONS

The proposed method has been implemented and evaluated using Python. The evaluation has been performed using various data sets available. The method has produced efficient results. The results produced have been presented below

Table 1: Details of Evaluation

Parameter	Value
Data Set	UPA
Number of Data points	5940
Number of Dimensions	53
Tool Used	Python

The details of evaluation has been used for the performance analysis of proposed algorithm has been presented in Table 1.

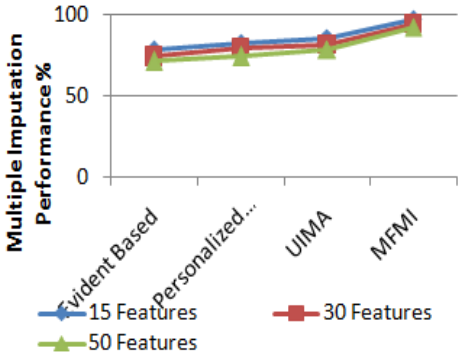

Figure 2: performance analysis on Multiple Imputation

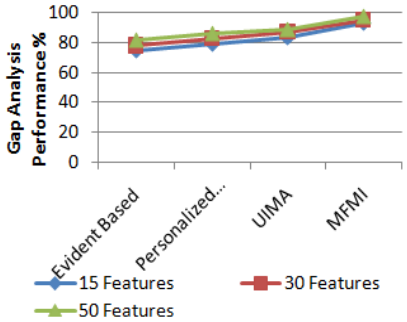

Figure 3: Comparison on GA Performance

The performance on multiple imputation performance has been measured and compared on varying number of features. The proposed MFMI algorithm has produced higher MI performance than other algorithms. The performance on gap analysis has been measured and compared with the results of other methods. The proposed MFMI algorithm has produced higher gap analysis performance.

V. CONCLUSION

The gap analysis in medical health care documents has been reviewed. Different approaches of gap analysis have been well studied in literature. From the literature it has been identified that, the result of gap analysis is based on the feature and technique considered. To improve the performance, a min feature approximation technique has been presented in this paper. The method reads the input data set and identifies the dimensions. For each dimension on each data point, the method fetches the feature value to compute the min feature approximation value. Based on the min feature value estimated, the method performs multiple imputations on identified missing values. The proposed method improves the performance of multiple imputation and gap analysis.

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. RhianSilvestro, "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; NestorasMathioudakis, 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. AnkeRohwerandAnelSchoonees,TarynYoung, 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. OlegMetskera, and EkaterinaBolgovaa, AlexeyYakovlevab, 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 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 using SAX Motifs, Random Forests and Text Mining techniques," 334–337, (2014).

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

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 qualitySpringerplus. 5,1 (2016).
18. BahaeSamhan ; K.D. Joshi, Resistance of Healthcare Information Technologies; Literature Review, Analysis, and Gaps, Hawaii International Conference on System Sciences,(2015).
19. K. Ananthajothi, M. Subramaniam, "Multi level incremental influence measure based classification of medical data for improved classification", Cluster Computing, The Journal of Networks, Software Tools and Applications, 2018. https://doi.org/10.1007/s10586-018-2498-z.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)