International Journal Of Core Engineering & Management

Volume-3, Issue-9, December-2016, ISSN No: 2348-9510

AN EFFICIENT EXTRACTION OF MEDICAL DOCUMENT AS SEQUENTIAL PATTERNS USING FUZZY LOGIC

M.Mallikarjuna Rao
Research Scholar
Rayalaseema University, Kurnool, Andhra Pradesh, India
mmkrao19@gmail.com

Dr.K.Fayaz
SKIM, Sri Krishnadevaraya University, Anantapuramu, Andhra Pradesh, India
fayazk16@gmail.com

Abstract

The classification of data and classify them to put into valid clusters is a challenge with more attributed data. Past works are on pre-processing with existing unstructured data and arrange them with various available techniques to segregation. This work proposed a new work SGM (System Generated Model) [3] to arrange the medical based thyroid live data to cluster [13][15] in an innovative approach. First data will be read and arranged with Meta data and later based on metadata SGM will take a fuzzy [4][1] decision to cluster the data further. MDM (Meta Data Model) will take the incoming valid checked pattern and classifies extracts it's metadata and once the index of the metadata framed. Based on the index and patterns SGM will frame documents [13][14] based on the input search criteria with respect to users sessions. This is incremental growth model which will be monitored by central repository for duplicate reduction to give feasible RM restuls from auto grown matrix by central repository before MDM to flush the existing output clusters. Fresh session follows MDM followed by SGM The intelligent validation check and clusters [16][12] is framing with monitor based on big data(thyroid medical data).

(Index items: fuzzy, Meta data model, system generated model, pattern, documents, central repository)

I. INTRODUCTION

This work is totally depends on medical based thyroid data with more than 1500 tuples framed up as live growing data with the addition of new decease with new patients data and their continuous updates.So once crawl starts for the valid combinational attributed data with this huge data, processing task is divided into two tasks by central repository. First classify **[15][12]** the data with validation and based on the preceding items patterns will be framed based on input criteria on selected attributes. Based on patterns **[13]** the Meta data from the base and loose data sets will be framed and based on metadata documents **[12][14]**framed with feasible time complexity.

International Journal Of Core Engineering & Management

Volume-3, Issue-9, December-2016, ISSN No: 2348-9510

II. AIM

The main aim of this work to reach to utilize less time to frame the flexible documents with heterogeneous combination of attributes. Based on the incremental growth of the live data feasible outputs and time consumption model will be balanced with reduction of repetitive searches **[5] [3]**. With the incremental growth of search and process with respect to duplication reduction and document **[15] [12]** framing time complexity has to be decreased with increase of accuracy.

Overall flow:

fig1:

III. METADATA

which gives micro information about the patterns with respect to central data with valid indexes as key values and the micro information (number of clusters and data offset values and address of the pattern with respect to cluster) of those particular patterns. Based on the key values and offsets**[9]**. The dynamically framed combinations will be shared into the clusters**[16][12]**. This metadata is framed up based on the priorities of the attributes. Total available attributes are 11 and central monitoring will be generating the prioritiesdynamically for each and every type of search. The available attributes are follows

fig 2:

The priorities based on the search criteria. Basically central repository assigns one fixed priority with the available attributes. But based on the search model the patterns will be framed**[3][12]** with fixed and variant priority set combinations **[13].**

International Journal Of Core Engineering & Management

Volume-3, Issue-9, December-2016, ISSN No: 2348-9510

IV. PRIORITIES BY CENTRAL REPOSITORY

ATTRIBUTE	PRIORITY
ADHAR NUMBER	1
PATIENT NAME	2
AGE	3
GENDER	4
COMPLICATIONS	5
DOCTOR NAME	6
DEPARTMENT	7
DOD	8
LAB REPORT	9
BLOOD GROUP	10

fig3:

If the search criteria is initiated and the input sequence with validation and followed by the cleansing model resides at central repository. The cleansing model checks with the patterns and compares the attributed set and cleans the non-available**[8]** items.

Example:

fig4:

The HEIGHT attribute is not the part of this set of attributes. Cleansing model will omit this items.

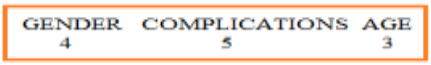


fig 5:

So now the priorities sequence by central repository is like this

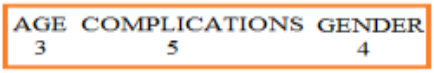


fig 6:

Compare to priority top priority of AGE(3) followed by GENDER(4) and COMPLICATIONS(5), But the input criteria is GENDER COMPLICATIONS AND AGE. So taking highest priority item as the most significant item for this patterns and checks**[3]** the

International Journal Of Core Engineering & Management

Volume-3, Issue-9, December-2016, ISSN No: 2348-9510

right side preceeding items. Since here there is no preceeding item to right and the very next left item is second significant item in this patterns and rest sequence will not be changed**[4]**. So the result is AGE(3) COMPLICATIONS(5) GENDER(4).

V. REDUCTION MODEL(RM)

Normally input and search based criteria can be repeated ,leads to time consumption of process**[2][4]** on the same inputs. This will be treated as one session with central repository and if the session repeats by comparing with central repository if duplication is available with LOG(maintains by central repository) automatically that process will be resulted from the central repository's LOG. This log will be updated continuously with each every session entry. This reduction model is an growing matrix model**[13][3]**
which is maintained and shared across to session boundaries but not with fresh sessions
Example model for RM:

SC1 SC2 SC3 are the search criteria available with LOG.

So each and every session contains one pattern and these patterns are available with the matrix format with central repository as LOG(past sessions):

SC1: ADHARNUMBER COMPLICATIONS DOCTORSNAME
SC2: AGE COMPLICATIONS GENDER
SC3: PATIENTNAME LABREPORT

	SC1	SC2	SC3
SC1	N	Y	Y
SC2	Y	N	Y
SC3	Y	Y	N

fig 7:

By taking the above table with the available sessions already with central repository and if new session(search criteria session LOG will be updated with central repository and if frameworks finds the duplication session then the matrix**[8][6]** will be in active to process further but the result which is with central repository log will be flushed as document clusters**[13][15]**.
Each and every session will be treated as its own process and these processes will reduce the time complexity by avoiding the duplicate processes**[4][1]** of flow.
Now the MDM will take the first item as AGE and starts with attributed value. Here the stack will be framed up with mapping model by collecting the meta data from the existing seed**[7]** or loose data.

International Journal Of Core Engineering & Management

Volume-3, Issue-9, December-2016, ISSN No: 2348-9510

VI. SAMPLE METADATA FORMAT

AGE COMPLICATION GENDER is the pattern and followed by metadata**[3][1]** format with 100 tuples

	AGE	COMPLICATION	GENDER
Indexes	34	78	67
Relative indexes	34	38	43
Super relative indexes	34	34	34

fig7:

VII. ALGORITHM FOR META DATA MODEL

Input: *Loose data sets with k attributes on n number of datasets*
Output: *Meta data structure to describe documents to* ***SGM*** *frame work*

Declaration Part:
|K| := Users input as sequential attributes
C[A] := Available attributes as vector
N = number of attributes available
∑Tm ← 0 //initialize attributed input
∑Fn ← 0 // Actual attributed input pattern
L{} **← 0** // Actual input pattern
Dm **← 0** // meta data information

MDM:
fc = COUNT(K) //assign input pattern sequence count to fc
L = K∈C //checks input patterns available in main
Count(C) = fc − SIZE(L) // actual count of input pattern items
for iteration 1 : n
START:
G(Dm[f] ,Dm[n] = META_DATA(L)
END;

International Journal Of Core Engineering & Management

Volume-3, Issue-9, December-2016, ISSN No: 2348-9510

VIII. SYSTEM GENERATED MODEL (SGM)

This is innovative approach to generated the clustered documents**[16][13]** instantly with respect to time complexity model. This acts with fuzzy model**[3][1]** box to generate the documents based on the meta information so that end result can be with feasible to crawl towards the micro points**[10][8]**. The main flow of this SGM is depends on meta data. Once the meta index framed SGM will take fuzzy**[3][2]** points generations over the data and frames the relative documents**[17][12]**
instantly. Each document framed up with nearest neibours of indexes and with average clubs with respect to second attribute priory.
Example: if the pattern input is **AGE COMPLICATIONS GENDER** and the documents will be framed based on AGE and preceding items(COMPLICATIONS) priority with respect to indexes. And it updates this to central log as invert matrix as invert matrix**[9][10]**.

As the result of the Meta data the information of AGE indexes available are 34 with COMPLICATIONS and GENDERS combinations. Here the documents will be framed**[3]** by putting marks at randomly generated points and fetches nearest neibours based on the second item COMPLICATIONS index and that block will be framed as clusters**[15].** The markers are will be different for next process if the input criteria are **AGE GENDER COMPLICATIONS** combination. Because the session could not be duplicated and this patterns is different from the previous.

```
Input: Meta data structure(Dm)
Output: grained relative documents (clusters)
Declaration:
M[ ] ← 0     // number of markers
n ← 0        // number of documents(clusters)
ΣITEM ← Dm   // assigning meta data structure to ITEM vector
I[] ← Item Vector  //temp storage of item vector for central repository
PLACEMENTS[]  = 0  // placements array to put markers
M[] ∃ MARK()
For each j in ITEM
START:
I[j] ← Dm[j]
PLACEMENTS[j] = J
END
CLUSTER_MODEL(I , PLACEMENTS, M)
```

This is the document pattern of the output which will produced by SGM and all the documents will be framed but with mapped with meta model based on the structure and will be only presented with mapping model but not physically. This is intelligent**[2][4]** model with fuzzy**[3]** technique with SGM.

International Journal Of Core Engineering & Management

Volume-3, Issue-9, December-2016, ISSN No: 2348-9510

After generation of placements each and every placements with fuzzy model**[4][2]** they will be with mark with results of the tuples(34) which is the count in the figure 7. As these marked places based on the count of the result(34) in this practical example case. If fuzzy**[3][2]** generates 3 then markers**[6][10]** would be 3 among 34 tuples based on the markers nearest tuple with respect to priority based documents**[13]**framed as follows:

Output results:

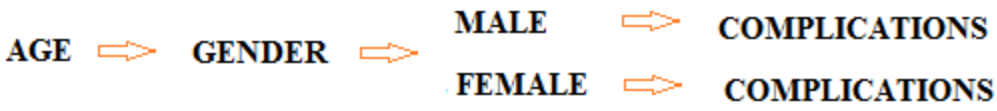


fig 8:

framed documents:

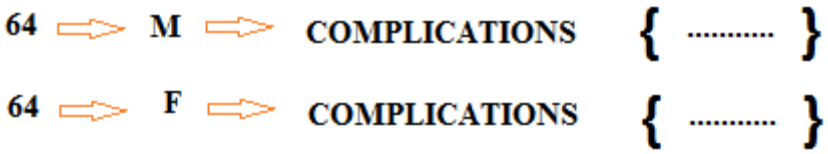


fig 9:

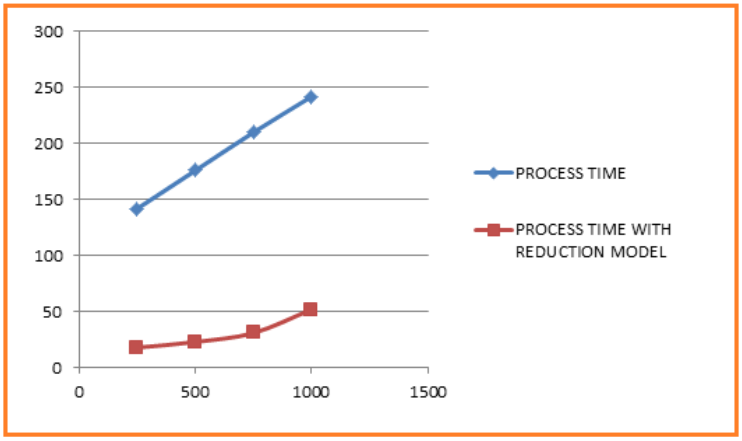
IX. RESULT ANALYSIS

With the above approaches by taking the pattern [AGE GENDER COMPLICATIONS] and after checking the results practically the following is the time consumption table based on tuples size(loose data) , process time and the process time with reduction model**[3]** based on the session.

TUPLIES SIZE	PROCESS TIME	PROCESS TIMEWITH REDUCTION MODEL
250	142	18
500	176	23
750	210	31
1000	242	52

International Journal Of Core Engineering & Management

Volume-3, Issue-9, December-2016, ISSN No: 2348-9510



With this work applying much number iteration always time complexity is totally feasible in all the combinational cases in all the aspects of documents**[14][12]** and their metadata**[13]** clarity.

X. FUTURE WORK

This work can be extended to **enhance the medical firm growth with respect to sessions growth of information extraction. The main attributes can be used for this growth could be complicated pattern analysis, frequency of sessions and nature of the users. Users(domestic and outsourced) behavioural analysis can be scaled and predicted with the patterns which the sessions and individuals framed(existing and new).**

REFERENCES

[1] Chung, S.K., Park, I.Y., Jang, T.Y., Chae, Y.M. Decision making support system in Otolaryngology part-2 (diagnosis of hearing loss). Korean J. Otolaryngol. 1989;32:768–789.
[2] Adlassnig, K.P., Kolarz, G., Scheithauser, W., Effenberger, H., Grabner, G. CADIAG—approaches to computer-assisted medical diagnosis. Comput. Biol. Med. 1985;15:315–333
[3] Nguyen, H.P., Torao, Y., Vu, Q.M. Approach to developing a diagnostic system combining diseases diagnosis of Western medicine with syndrome differentiation of traditional [4] medicine. J. Biomed. Soft. Comput. Hum. Sci. 2000;5:9–16.
[4] Nguyen, H.P. Towards Intelligent Systems for Integrated Western and Eastern Medicine. TheGioi, Hanoi; 1997.
[5] Haux R. Medical informatics: past, present, future. Int J Med Inform. 2010; 79(9): 599-610

[6] Guarino N. Formal Ontology in Information Systems. Proc FOIS. 1998; 3-15
[7] Marwede D, Fielding M, Kahn T. RadiO: A prototype application ontology for radiology reporting tasks.Proc AMIA. 2007; 513-517
[8] Schober D, Boeker M, Bullenkamp J, Huszka C, Depraetere K, Teodoro D, Nadah N, Choquet R, Daniel
C, Schulz S. The DebugIT core ontology: semantic integration of antibiotics resistance patterns., Stud
Health Technol Inform. 2010; 160(Pt 2): 1060-1064
[9] Maimon OZ, Rokach L. Data Mining and Knowledge Discovery Handbook. NY: Springer; 2005
[10] Iakovidis DK, Tsevas S, Savelonas MA, Papamichalis G. Image Analysis Framework for InfectionMonitoring. IEEE Trans Biomed Eng. 2011; 59(4): 1135-1144.
[11] Mitchell HB. Multi-sensor data fusion. Springer-Verlag; 2007
[12] Ran Vijay Singh and M.P.S Bhatia, "Data Clustering with Modified K-meansAlgorithm", IEEE International Conference on Recent Trends in Information Technology,ICRTIT 2011, pp 717-721.
[13] D. Napoleon and P. Ganga lakshmi, "An Efficient K-Means Clustering Algorithm forReducing Time Complexity using Uniform Distribution Data Points", IEEE 2010.
[14] Tajunisha and Saravanan, "Performance Analysis of k-means with differentinitialization methods for high dimensional data" International Journal of ArtificialIntelligence & Applications (IJAIA), Vol.1, No.4, October 2010
[15] Neha Aggarwal and Kriti Aggarwal,"A Mid- point based k –mean ClusteringAlgorithm for Data Mining". International Journal on Computer Science and Engineering(IJCSE) 2012.
[16] Barileé Barisi Baridam," More work on k-means Clustering algortithm: TheDimensionality Problem ". International Journal of Computer Applications (0975 – 8887)Volume 44– No.2, April 2012.
[17] Kohei Arai,Ali Ridho Barakbah, Hierarchical K-means: an algorithm for centroidsinitialization for k-Means. Reports of the faculty of Science and Engineering, SagaUniversity, Vol. 26, No. 1, 2007.