Multi Aggregator for Content Aggregation and Contextual Learning

Sana Maqsoon Khan[#1]

[#] *CSE Department, TIT College, RGPV University*
Bhopal, India
[1] *sanamaqsood786@gmail.com*

Abstract – Content aggregation and multi-media data processing over the aggregated data is challenging research to provide efficient mechanism for better experience. Content aggregator contains multiple source data and in different format which can store and further process, use by the vendor. Here the problem associate is data acquisition, data storage and further processing through. In our dissertation the work is taken from different content mining, image processing and relevant matching over the content. A technique of acquisition is also going to describe which make use of advance technique and provide a proper content mining and output to the user. In existing technique various model such as HMM (hidden markov model), co-occurrence model and various cross validation being presented. Thus the approach use by them given as further work in the field of data acquisition and applying more efficient parameter to compute the efficiency. In our dissertation a work on determining the approach for a data acquisition and further content processing by the relation of multiple content aggregators is used. Our approach make use of enhance parameter computation which shows the efficiency of our outcome. Experiment were performed via CPU machine 4 GB RAM, 1 TB Hard disk and i5 processor. It found satisfactory which is implemented using Java language with file media access API. Thus the observed parameter such as precision, recall, Accuracy found to be better while comparing with previous algorithm of working with single aggregator system.

Index Terms - **Content Aggregator, Media access, Data acquisition, Content processing, data observation, Confusion matrix, parameter evaluation.**

I INTRODUCTION

In current scenario a huge amount of data is introduced by the various internet applications. User uses these applications and generates data, storage for that data which is in heterogeneous form and scattered in the space. Thus a centralized mechanism is required which Provide a centralize storage and access to that data. There are various fields like social media sites, digital equipment, medical science, military intelligence etc. that data contains content from these fields and stored at the server, so to access that it's a time consuming process to analyze that data and provide desired data to the user, a centralized data management system which provides a centralize functionality to the user to access that data should be needed.
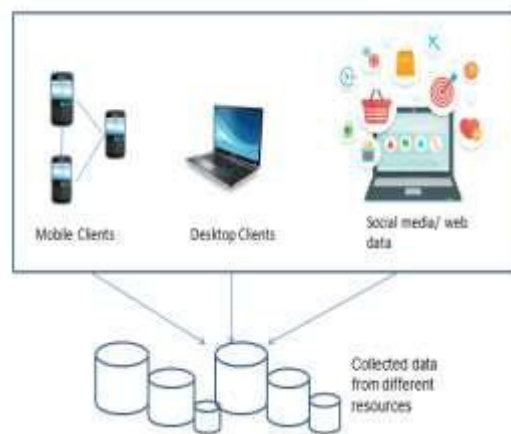

Figure 1.1: Distributed content in the space.

The above diagram shows that there are lots of information all over the world and all the information is coming from different resources like from mobile data clients, social media data in the form of text and images and videos as well. All the data comes from different resources and save in the database.

An aggregator is a collection of the data which contains data from different fields and provide a centralized access for that data. In existing techniques a single aggregator is used to do that task but that not be able provide desired performance to search that data. Thus this dissertation is a new techniques which first divide data into different fields and then apply that online distribution based method to aggregate that data. That provides an enhanced framework to access that data and also increases the search efficiency.

II Content Aggregator

Aggregators are either a website or computer software which provides aggregation of the data on the basis of the features and information. That helps to suggest data to the users for their searches. There are various types of aggregators which used to aggregate data from different fields or categories, a aggregator is shown in Figure 1.2, that shows aggregation of the data of

different fields like business, marketing, online shopping etc.

Figure 1.2: Content Aggregator.

The above diagram shows that here is a content aggregator, which collects overall information that is coming from resources like business data, shopping data.

There are various types of aggregators like:

Data Aggregator –
In that data is collected from the databases or storage, then datasets are generated to provide as a input for various data processing applications or data analysis, like in public records or criminal databases. That aggregated datasets provides an enhanced functionality for the data processing.

Video Aggregator [1] [2] - Video aggregators are the aggregators which are used to organize videos from the different online sites and provide a centralize system to access these videos. There are many video aggregators like youtube, videosit, dailymotion etc, that used to provide videos to the user.

Search Aggregator [3]- search aggregator is the type of search engine which contains meta-search data which combine search from various search engines. It uses user's previous search feeds to provide control over the content by the aggregator.

Figure 1.3: A description for feed Aggregation.

The above diagram is of description for feed and aggregation, feed data is coming from the facebook, youtube, weather data, news etc.

Social Network Aggregations-
It is the process of collecting content from various social network sites like youtube, facebook, instagram etc to provide an aggregation for the social media content into a single presentation.
So there are various types of applications like in e-commerce sites, in video sites and some other. In Figure 1.2, a aggregator with different types of feeds which contains details about various types of sources is shown. That provides an centralize system to access for that data.

III Objective

To search multimedia content is a difficult task to do because that data presents in scattered form and poses heterogeneity in nature, thus it's hard to search that data. There are various type of recommendation systems are provided by the various researcher to provide better search experience to the user. Thus a new multiple aggregator based system is presented in this dissertation. That first divides all the data in to different categories and then a aggregation is applied to generate aggregator for different techniques. These aggregators provide a better recommendation for the user in context their search.

Summery
Content aggregator is the process of collecting data from various sources which contains useful information and provide a centralize system to access that data. An overview over the various types of content aggregator and their application area where these techniques are used is presented in this chapter.
IV Literature Survey

A. Introduction
There various techniques are presented by the researchers to provide an aggregator. A review over the various techniques which used to provide a content aggregator is presented in this chapter. There are
various techniques like collaborative filtering,distributed online learning, content based clustering etc, technique are used to aggregate content.

B. Literature Review

CemTekin, Mehaelavan Der Schaar
[1], the huge growth in multimedia content in last decade. That data generated by multiple sources which is generated in heterogeneous form because data is generated by different resources that generate variety in the data which make difficult for a user to get on demand data. in this paper a distributed online matching based content aggregator is present which gathers data from different resources and suggest user data in context of their search. A distributive online matching algorithm is used to match the content which

generated by heterogeneous producers in that algorithm aggregator can learns from the experience of the other CA's, in that CA can suggest content from multimedia sources which

is that not connected directly to it. It learns about the content from the CA which is connected to it.

Eric Bruno S

[2] Inthis paper a multimodel preference based aggregation technique is presented in which joint design a multimodel representation and a machine learning based fusion algorithm is used in which first a preference based representation is defined and then a rankboost fusion algorithm is used to map user's query.

Preference space used in this paper is degraded but lightweight representation of original spaces is achieved, that model does not have such issues like dimensionality and space thus it can be able to fusion heterogeneous data. Rankboost algorithm facilitates to match the functionality of the whole multimodel system and provide a fast fusion mechanism for heterogeneous data.

ReolandOrdelman, Fransiciska De Jong, Martha Larson

[3] In this paper semantic speech retrieval based technique is presented which provide a way to access multimedia data. This technique focuses on automatic speech recognition and spoken document retrieval, this technique is used for multimedia access in that broadcast news data puts a huge advantage and in that case that is in clean and organized structure which is used for but this technique still not used so widely because uncertainty in the speech data, this technique uses out of vocabulary words to enhance the performance of the technique and in that way it is an efficient technique but it need some improvement to use in real world content retrieval.

Summery
A review over the various techniques which is used to provides content aggregation for various applications like data analysis, to analyze data an aggregated dataset is used, Searching, to perform any search operation a aggregated data will be required, E-commerce, to recommend products to the user a aggregated data is needed. There are techniques like collaborative filtering, distributed online learning and some other techniques are used.

V. Problem Identification & Methodology

As per monitored previous work using which the content aggregation, data acquisition and content retrieving is performed. Here are some issues which arrived after review the work in detail.

1. The existing technique make use of hardware device and complex level coding such that it's not feasible in every section at micro level.
2. A single aggregator and availability can be a lack of resource or limited resource on which the working need to get perform.
3. Arecent content based technique face the problem such as cold start, vulnerable for attacks in the dataset and processing technique which is still be an open challenge to work on.
4. A security mechanism along with large level data processing is not been implemented properly which can use in real time with various resources and outcomes.
5. Existing technique make use of digital media where the acquisition is not been described and thus no dynamic significance is available.
6. Content aggregator suggestion and filter pre-process is not given in most of the previous available techniques.
7. Traditional technique minutia based, triangle or color based technique puts their limitation in detection of scene.
8. Existing techniques having low computation parameter such as precision, recall, accuracy.

Thus the further work is required in data acquisition and data processing over digital media content and aggregation.

VI. Proposed solution

A. Introduction
Content aggregator is used to collect information from the various resources and provide that information to the user. That information plays a vital role in various applications to improve the performance. In existing techniques a single aggregator is used to do all the tasks of aggregating data from various sources , that data contains data which poses different properties.

There are various techniques like collaborative filtering, content based extraction, distributed online learning.

Collaborative Filtering:

In collaborative filtering used in two prospective, one general and other one is narrow. In general, multiple agent based technique is used to match patterns filtering information on the basis of collaboration among these agents. In narrow prospective, user's previous behaviors are used to recommend content for the user.
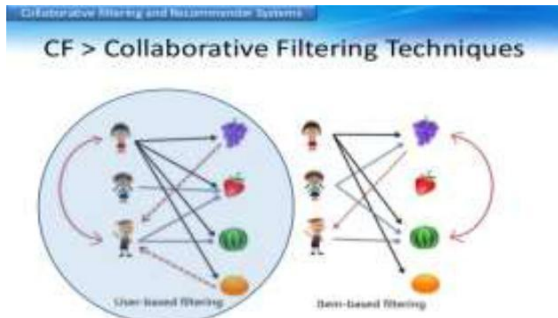
Figure 4.1: Collaborative filtering

Content based filtering:

In content based filtering technique is used to fetch items which poses similar content, and predict items to the

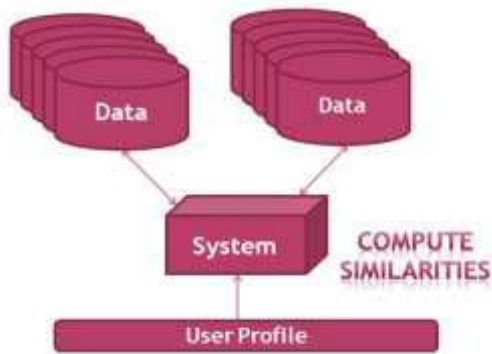Various users on the basis of correlation of user's previous behavior.

Figure 4.2: Content Based Filtering.

The above diagram shows content based monitoring, on the basis of information and characteristics of the items. All the data and user

profile data will go to the system and compute the similarities.

VII. Distributed online learning:

In distributed online learning, data from the various nodes of the database is collected and that result is provided to the coordinator. The query coordinator combines these results used to produce a concrete result, that coordinator continue that process further to refine the aggregations.
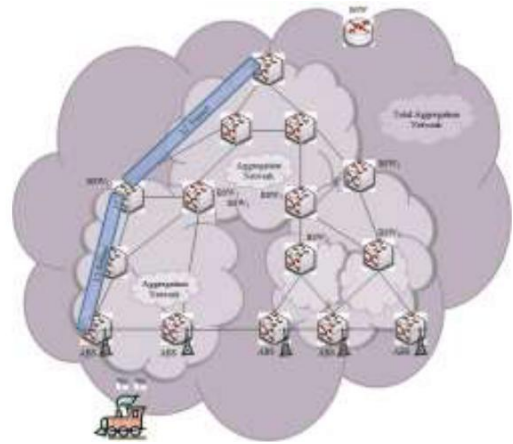

Figure 4.3: Distributed online learning

VIII. Proposed Technique

In existing a enhance technique is presented by the []. In that technique a CA to CA learning process is presented. In that a CA (content Aggregator) can read data from other aggregator rather than the main source. That reduces the time to access data. But in that technique a single aggregator is used which is not enough to provide desire performance as required. Thus a new enhanced technique a multiple content aggregator based technique is presented. In that technique firstly data from the sources is divided into different categories and then online learning technique is applied to form clusters. A flow-diagram and algorithm for the proposed technique is presented below.
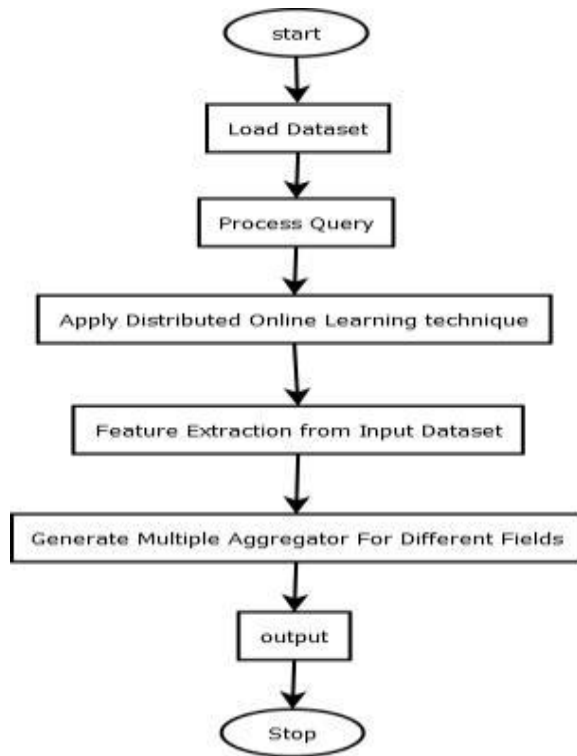
Proposed technique flow diagram

Figure 4.4: Flow-Diagram for Proposed Technique

Algorithm for proposed technique is as follows:

Input: Query Image, Dataset, CA.
Output: Query Output.
Steps:
  Start
  Load complete dataset (i= 0-n)
for (int i=0;n)
  {
  Load (i);
  }
for (int i=0;n)
  {
  Process query using distributed online learning,
Generate aggregator for different fields. }

for (int i=0;n)
  {
Feature extraction from the all input data sets
Return: output
}
End;

Summery
A description over the proposed technique is presented in this chapter. There are various techniques like Collaborative filtering, Content based filtering etc, are used to aggregate content and provide aggregator to search content. Online distributed learning based

technique is used which provides an enhanced aggregation for the data. In that a multiple aggregation based technique is presented, which provide an enhanced functionality to access content.

IX. Implementation

A.Introduction
Content aggregator provides an easy way to search contentent because it provides an easy and centralized access for the data which resides in different resources. A description over the proposed technique is presented in this chapter. In that, To implement proposed technique a JAVA language over NETBEANS IDEis used. A result analysis for the proposeed techique is also presented in this chapter which shows that proposed techique provides an enhanced funtionality to search.

B. Implementation Details
A detailed description about the implementation details is presented in this section. To implement proposed techique a java language over NETBEANS IDE is used. NETBEANS IDE is an open source simulator which contains inbuilt tools to implent project in java had some other languages in that way it provides an user friendly environment to develop program

X.Result and Analysis
To estimate the performance of the system, the following formulas are used.
Classification rate = (Number of classified patterns * 100)/ Total number of patterns. True positive rate = TP/TP+FN
False positive rate = FP/FP+TN

On applying probability based feature selection algorithm, following features were selected.

- There are result parameter to monitor the efficiency of that algorithm is as follows[5,6]-

- Precision – this parameter value can be calculated by processing the algorithm, precision is number of retrieval data to the query..

- Precision = (relevant data ^ retrieved data)/retrieved data.

- Recall – this parameter value can be calculated by processing the algorithm, Recall in information retrieval is the fraction of the documents that are relevant to the query that are successfully retrieved.
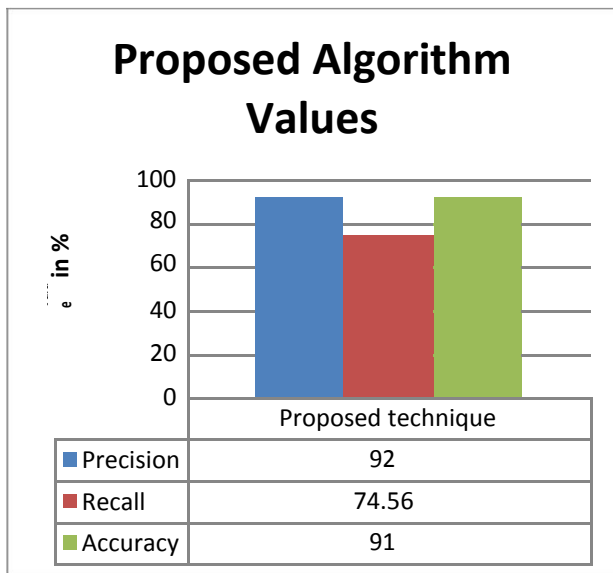
Recall = (relevant data ^ retrieved data)/relevant data
As per the observed result and experiment setup [7], technique is implemented. The proposed and existing technique is performed with the above dataset presented and the algorithm performed with the system and following output results was monitored.

A. Data result comparison

The above table represents the number of data values from the data and algorithm is performed.

Existing Algorithm Values

	Existing Technique
■ Precision	82
■ Recall	78
■ Accuracy	87
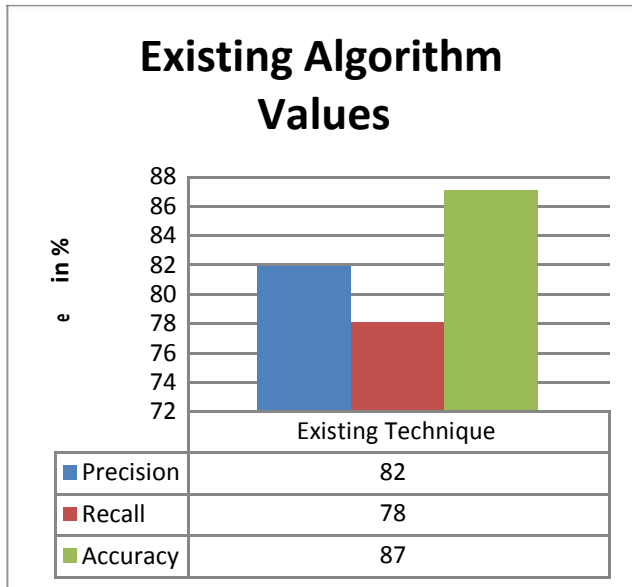
Graph 1: Existing Algorithm parameter values

Proposed Algorithm Values

	Proposed technique
■ Precision	92
■ Recall	74.56
■ Accuracy	91

Graph 2: Proposed approach parameters

The above graph no.1& graph no.2 represent the number of data values from the data and algorithm is performed.

Comparison graph

	Existing	Proposed algorithm
■ Precision	82	92
■ Recall	78	74.56
■ Accuracy	87	91
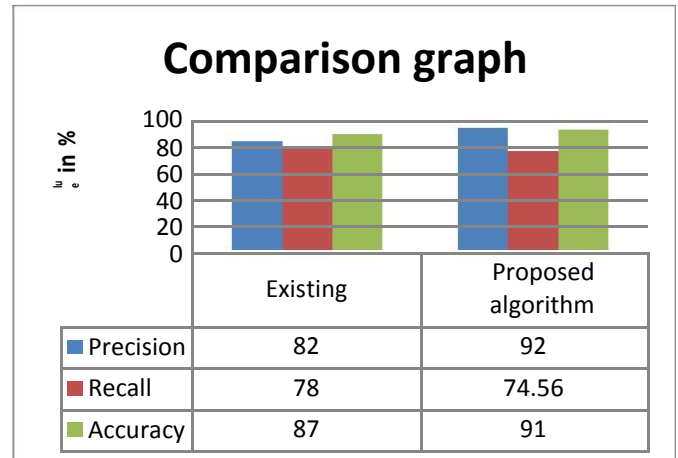
Graph 3: Comparison graph for technique analysis x axis in % parameter value.

The Above graph demonstrate the complete picture of comparison in between the available technique and our proposed work technique , thus further the execution outperform its best while comparing with the existing work approach.

XI.Conclusion and Future Work

Content aggregation is the process of collecting content from the various resources which contains useful information for the same topic. There are various application fields like
E-commerce, data processing, where content aggregation can be used. It provides an easy and centralized access for the data from various resources. There are various types of aggregators like data aggregator, video aggregator, poll aggregator, etc, which can be used for different applications.

There are techniques like content based filtering, collaborative filtering, online distributed learning etc are used to aggregate content and generate a Contentaggregator for the various application's description over thesetechnique is presented in chapter 3.

In existing technique a single aggregator based technique is used to provide aggregation. In that a distributed online learning technique can be used to aggregate content. But a single aggregator is not enough to aggregate that content. Thus a multiple aggregator based technique is presented in this dissertation, in this technique first data is divided into different fields and then a distributed online learning based technique can be used to provide aggregator for these fields. A CA to CA learning technique also used in that a content aggregator can be learn from the other content aggregator rather resources that provide an efficient mechanism to search.

A.Future Work

For future work a new and enhanced aggregator technique can be used to provide better functionality to aggregate content and provides an efficient and cost effective way to recommend content to the user in context of their search.

REFERENCES

[1] Cem Tekin, Mehaela van Der Schaar "contextual online learning for multimedia content aggregation" IEEE Transactions on multimedia Vol. 17 No. 4, April 2015.

[2] Eric Bruno, StephaneMarchandMailet" multimodelpreference aggregation for multimedia content aggregation" journal of multimedia vol. 4 No.5, October 2009.

[3] ReolandOrdelman, Fransiciska De Jong, Martha Larson "Enhanced Multimedia content access and Exploitation using semantic speech retrieval" IEEE conference on Semantic Computing, 2009.

[4] Alberto Messina, Maurizio Montagnuolo "A Generalised cross-model clustering method applied to multimedia news semantic indexing and retrieval " April 2009.

[5] Sameer Amir, PetrikBlandin, Laon Marius Bilasco, Chabane Djeraba" semantics for intteligent delivery of multimedia content" 2010.

[6] Kenneth K. Fletcher and Xiaoqing (Frank) Liu "A Collaborative Filtering Method for Personalized Preference-based Service Recommendation" IEEE, 2015.

[7] Chih-Lun Liao, Yu-Chun Lin, Shing-Tai Pan, Shie-Jue Lee "Improving Efciency of Recommender Systems" IEEE, 2015.

[8] T. Veugen, Z. Erkin "Content-Based Recommendations with Approximate Integer Division" IEEE, 2015.

[9] Wu wenjuan, Lu zhubing "A Personalized Recommendation Strategy Based on Trusted Social Community" IEEE, 2015.

[10] Haifeng Liu, Xiangjie Kong, Xiaomei Bai, Wei Wang, Teshome Megersa Bekele, Feng Xia "Context-Based Collaborative Filtering for Citation Recommendation" IEEE, 2015.

[11] Chia-Chi Wu, Yi-Chong Zeng, and Meng-Jung Shih "Enhancing Retailer Marketing With An Facial Recognition Integrated Recommender System" IEEE, 2015.

[12] Cem Tekin, Member, Mihaela van der Schaar "Contextual Online Learning for Multimedia Content Aggregation" IEEE, 2015.