Dataretrieving for varied in different Composition Databases using Content aggregation

Mangesh S. Khode, Mayur S. Dhait
Computer Science and Engineering, RTMNU University, Agnihotri College of Engineering,
Nagthana Road, Sindi(Meghe),Wardha, Maharashtra, India
mangeshskhode@gmail.com

Abstract:-Keeping in mind with a variety of content choices, consumers are exhibiting diverse preferences for content; their preferences often depend on the context in which they consume content as well as various exogenous events. To satisfy the consumers' demand for such diverse content, multimedia content aggregators (CAs) haveemerged which gather content from numerous multimedia sources**.** A key challenge for such systems is to accurately predict whattype of content each of its consumers prefers in a certain context,and adapt these predictions to the evolving consumers preferences, contexts, and content characteristics This paper addressesgenerate text based file data sets, such as word, text files, image file data sets, and video file data sets, It also extract data from multiple databases, evaluate user preference based query, reduce time complexity by clustering data, and increase fetching speed by using query classification.

Keywords: *Content aggregators, data sets, query classification*

*****

1. Introduction

A quantity greater than needed of multimedia applications (web-basedTV personalized video retrieval personalizednews aggregation etc.) are emerging which requirematching multimedia content generated by distributed sourceswith consumers exhibiting different interests.

Multimedia gives personalized video retrieval, news aggregations. They are emerging which require matching multimedia content generated by distributed sources with consumers exhibiting different interests. The matching is often performed by CAs that is responsible for mining the content of numerous multimedia sources in search of finding content which is interesting for the users. Each user is characterized by its context, which is a real valued vector that provides information about the users' content preferences. An example of the system with users, CAs and multimedia sources is given in following Fig.1.

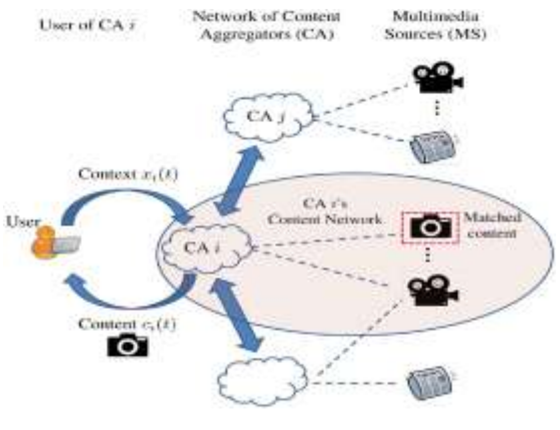


Fig. Operation of the distributed content aggregation system. (a) A user with type/context $x_i(t)$arrives to Content Aggregator (CA)i . (b) CA chooses a *matching action* [1]

We assume a model where users arrive sequentially to a CA, and based on the type (context) of the user, the CA requests content from either one of the multimedia sources that it is connected to or from another CA that it is connected to. The CA's role is to match its user with the most suitable content, which can be accomplished by requesting content from the most suitable multimedia source.1 Since both the content generated by the multimedia sources and the user's characteristics change over time, it is unknown to the CA which multimedia source to match with the user. This problem can be formulated as an online learning problem, where the CA learns the best matching by exploring matchings of users with different content providers. After a particular content matching is made, the user "consumes" the content, and provides feedback/rating, such as like or dislike. [1]

A general definition of context was proposed by Chen and Kotz: "Contextis the set of environmental states and settings that either determines an application's behavior or in which anapplication event occurs and is interesting to the user." Considering the IPTV service, context canbe considered as any information that can be used to characterize the situation of an entity related to the IPTV service. An entity could be the user, device, and network and service itself. Thus we define four types of contexts for IPTV chain including user, device/terminal, network and service domains. In order to enable context-aware IPTV for enriched services personalization, variety of information expressing current situation of user, device, network, content and service needs to be collected and processed. Such information is called contextual information and needs to be efficiently gathered and processed in real-time during service access. [3]

The term social multimedia to refer to multimedia resources available via social media channels or more formally: online sources of multimedia con- tent posted in settings that foster significant individual participation and that promote community duration, discussion and re-use of content. Social multimedia presents a significant opportunity for Multimedia applications and services. Such information may include many facets: textual descriptors, information about the location of the content capture the camera properties metadata, and even user information and social network data. These additional metadata can be used to advance and augment multimedia and content analysis techniques. In addition, social multimedia captures and leverages community activity around multimedia data, using explicit user input like tags and comments as well as implicit input from users like mass viewing patterns in item and sub-item levels. Indeed, social multimedia also offers the opportunity to design interactive systems that elicit new explicit and implicit metadata from user interaction. Such interaction and user input is often driven by social motivations and can improve the data available for multimedia applications. Thus, social multimedia offers several opportunities that go beyond and above other \Web multimedia" sources where many of these opportunities are not available. [7]

2. Content Aggregation Concerns

For implementing content aggregation of multiple heterogeneous databases, we use two basic algorithms and one technique:

2.1 K-means clustering algorithm for data classification for query based results

It is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining. K-means clustering aims to partition observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. This results in a partitioning of the data space into cells. The problem is computationally difficult (NP-hard); however, there are efficient heuristic algorithms that are commonly employed and converge quickly to a local optimum. These are usually similar to the expectation-maximization algorithm for mixtures of Gaussian distributions via an iterative refinement approach employed by both algorithms. Additionally, they both use cluster centers to model the data; however, k-means clustering tends to find clusters of comparable spatial extent, while the expectation-maximization mechanism allows clusters to have different shapes. The algorithm has nothing to do with and should not be confused with k-nearest neighbor, another popular machine learning technique. It helps in uploading of data over the server.

2.2 Distributed Content Aggregation Algorithm

- Content aggregation query
- Send request to all available search engines using query
- Collect data from individual servers and generate clusters using K-Means
- Forward result to clients

2.3Data extraction or fetching

Data extraction or fetching of data from multiple sources has touse query based search engine. It is done by content aggregation.

3. Literature Review

CemTekin and Mihaela van der Schaar [1] proposed a novel, distributed, online multimedia content aggregation framework, which gathers content generated by multiple heterogeneous producers to fulfill its consumers' demand for content. To satisfy the consumers' demand for such diverse content, multimedia content aggregators (CAs) have emerged which gather content from numerous multimedia sources. A key challenge for such systems is to accurately predict what type of content each of its consumers prefers in a certain context, and adapt these predictions to the evolving consumer's preferences, contexts, and content characteristics. We propose since both the multimedia content characteristics and the consumers' preferences and contexts are unknown, the optimal content aggregation strategy is unknown a priori. Our proposed content aggregation algorithm is able to learn online what content to gather and how to match content and users by exploiting similarities between consumer types.

MarekDabrowski, JustynaGromada, HassnaaMoustafay and Jacky Forestier [3] proposed pervasive computing and context-awareness principles seem to be promising for making user interaction with the system more seamless and fluid. A novel architecture for unified storage and processing situational data in IPTV service domain is presented, together with discussion of its implementation issues and validation by test bed experiments. TV and video services landscape is currently undergoing significant changes. Traditional TV broadcasting model is supplemented and often even replaced by digital content distribution services over the Internet. As a result of this trend, a range of services and contents available for users is rapidly expanding. As a side-effect, designing efficient user interfaces for discovering the content, as well as for manipulating associated interactive services, becomes more and more cumbersome. Pervasive computing and context-awareness principles seem to be promising for making user interaction with the system more seamless and fluid.

S. Roy, T. Mei, W. Zeng, and S. Li [6] proposed learning for prediction of video popularity socially. Cross domain real- time transfer learning framework is used which utilizes knowledge from social streams (e.g., Twitter) and improve popularity prediction in the video domain. OSLDA model is used to detect topics from social streams. Social Transfer algorithm is used for classifying videos with topics which is then used to calculate the social prominence and finally leading to the improved popularity prediction in the video domain. The framework has the ability to scale with incoming tweets in real time. Knowledge gained from social streams can be used to address many multimedia problems which cannot be solved by using traditional multimedia techniques alone.

PM. Naaman[7] proposed various Web-based sharing and community services such as Flickr and YouTube have made a vast and rapidly growing amount of multimedia content available online. This article presents an approach for "social multimedia" applications. The approach is based on the experience of building a number of successful applications that are based on mining multimedia content analysis in social multimedia context. In recent years, various Web-based sharing and community services such as Flickr and YouTube have made a vast and rapidly growing amount of multimedia content available online.

M. van der Schaar, J. Xu, and W. Zame [8] proposed and analyzed protocols that rely solely on the exchange of fiat money or tokens. The analysis has much in common with work on search models of money but the requirements of the environment also lead to many differences from previous analyses and some surprises; in particular, existence of equilibrium becomes a thorny problem and the optimal quantity of money is different. In many online systems, individuals provide services for each other; the recipient of the service obtains a benefit but the provider of the service incurs a cost.

L. Li, W. Chu, J. Langford, and R. E. Schapire [9] proposed personalized recommendation of news articles as a contextual bandit problem, a principled approach in which a learning algorithm sequentially selects articles to serve users based on contextual information about the users and articles, while simultaneously adapting its article-selection strategy based on user-click feedback to maximize total user clicks. The contributions of this work are three-fold. First, we propose a new, general contextual bandit algorithm that is computationally efficient and well motivated from learning theory. Second, we argue that any bandit algorithm can be reliably evaluated offline using previously recorded random traffic. Finally, using this offline evaluation method, we successfully applied our new algorithm to a Yahoo! Personalized web services strive to adapt their services (advertisements,
news articles, etc.) to individual users by making use of both content and user information. Despite a few recent advances, this problem remains challenging for at least two reasons. First, web service is featured with dynamically changing pools of content, rendering traditional collaborative filtering methods inapplicable. Second, the scale of most web services of practical interest calls for solutions that are both fast in learning and computation.

M. Saxena, U. Sharan, and S. Fahmy[10]proposed different DNS resolvers to obtain the IP address of the video server. We study how the DNS resolution impacts the performance of the video download, thus the video playback quality. As the tool is intended to run on multiple ISPs, we have discovered some interesting results in YouTube distribution policies. These results can be applied to any content delivery networks (CDN) architecture and should help users to better understand what the key performance factors of video streaming are. Online video services account for a major part ofbroadband traffic with streaming videos being one of the most popular video services. We focus on the user perceived quality of YouTube videos as it can serve as a general index for customer satisfaction.

G. Linden, B. Smith, and J. York[13] proposed recommendation algorithms to personalize the online store for each customer. The store radically changes based on customer interests, showing programming titles to a software engineer and baby toys to a new mother. The click-through and conversion rates two important measures of Web-based and email advertising effectiveness vastly exceed those of untargeted content such as banner advertisements and top-seller lists. There are three common approaches to solving the recommendation problem: traditional collaborativefiltering, cluster models, and search-based methods. Here, we compare these methods withour algorithm, which we call item-to-item collaborative filtering. Unlike traditional collaborativefiltering, our algorithm's online computation scales independently of the number of customers andnumber of items in the product catalog. Our algorithm produces recommendations in real-time,scales to massive data sets, and generates high quality recommendations.

D. Bouneffouf, A. Bouzeghoub, and A. L. Gançarski [14] proposed location and time context for examples are important in mobile computing recommendations, due to the fact that a user may require a recommendation at a particular location in a particular time. Such scenarios have introduced Context-Aware Recommender Systems (CARS) for further open research issues and challenges. This paper initially presents Background of CARS, specifically from the perspective of context types, context modeling architectures and algorithms. Furthermore, the paper, presents an overview of the state-of-the-art research in the area of CARS, and finally discusses relevant open issues of CARS.Recommender Systems have been/are being researched and deployed extensively in various disciplines such as tourism and education. Most traditional recommender systems such as Collaborative Filtering (CF) and Content-Based Filtering (CBF) generate recommendations by using two main attributes, namely; users and items i.e. recommendations are generated based on a user having an interest or preference of a particular item resource.

E. Hazan and N. Megiddo[15]proposed the framework by allowing an experts algorithm to rely on state information, namely, partial information about the cost function, which is revealed to the decision maker before the latter chooses an action. This extension is very natural in prediction problems. For illustration, an expert's algorithm, which is supposed to predict whether the next day will be rainy, can be extended to predicting the same given the current temperature. We introduce new algorithms, which attain optimal performance in the new framework, and apply to more general settings than variants of regression that have been considered in the statistics literature. The standard so-called experts algorithms are methods for utilizing a given set of "experts" to make good choices in a sequential decision-making problem. In the standard setting of expert's algorithms, the decision maker chooses repeatedly in the same "state" based on information about how the different experts would have performed if chosen to be followed. In this paper we seek to extend this framework by introducing state information.

 ISSN: 2454-4248
34 - 40

4. Methodology and System Model

1. Admin Module:
Generating dataset for links
Generating dataset for images
Generating dataset for videos

2. Client Module:
While searching data will be fetched from all web servers using content aggregation for datasets.

3. Architecture and Flow Diagram:

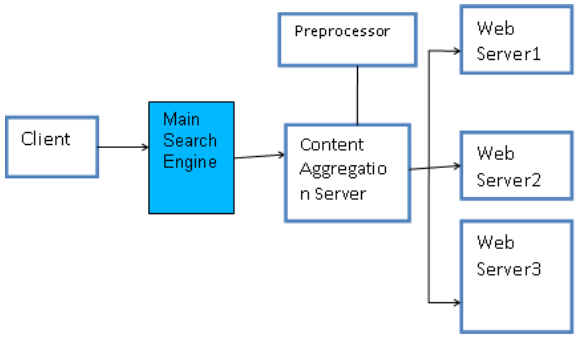


Fig. System Model

5. Proposed System
The proposed work is planned to carry out in the following manner.

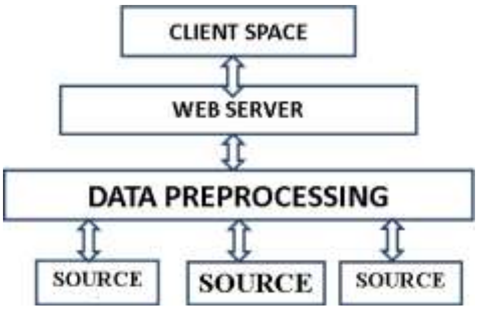


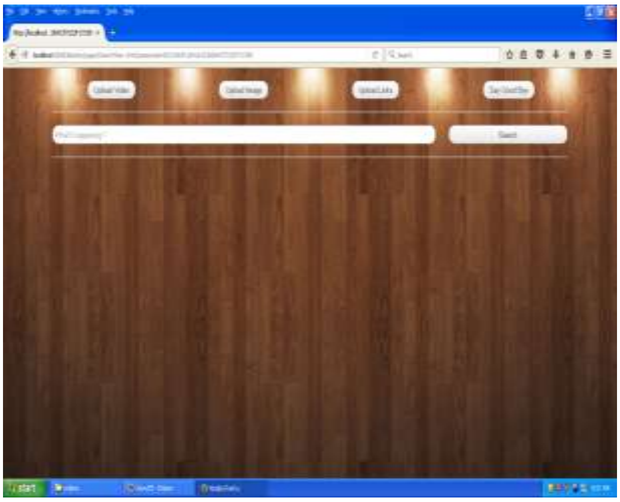
Fig.2. System Architecture

In this proposed system, then client from client space will be able to fetch data from multiple databases which contain different media. The user will be provided with search engine to search text, image or video based files. The proposed system will provide functionality of distributed system in virtual centralized view approaches shown in above figure.

We will be using three different databases or folder for separate search. It removes ambiguity of sources having different web servers by providing common web server. For that, data preprocessing plays vital role.

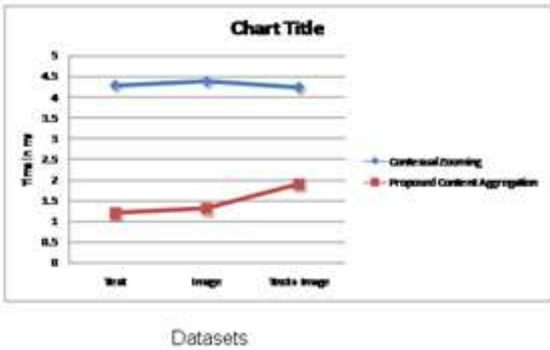
6. Implementation
We implement our main task in tomcat web server. We execute 3 servers namely image, text and video separately along with Google search server.

ISSN: 2454-4248
34 - 40

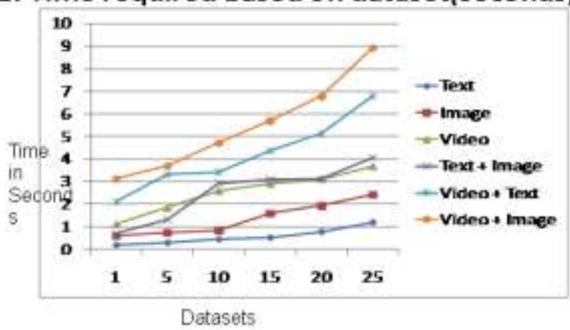


7. Results and Discussion

• 1. Time to fetch data from single server(



• 2. Time required based on dataset(seconds)



8. Conclusion

Content aggregation is one of the most important requirement for data search engines. To efficiently fetch data from multiple web servers, there is a requirement of fast data collection server which can collect, preprocess data and sent to another clients. In proposed work, we have implemented a content aggregation service which can fetch data from multiple web search engines and combine results which can be provided to clients. Through results we can verify that algorithm works efficiently in distributed domain.

9. Future Scope

To improve results using better recommendation service. Implement audio based search engine. Implement document search engine. Aggregate much more search engines efficiently.

REFERENCES

[1] CemTekin, Member, IEEE, and Mihaela van der Schaar, Fellow, IEEE. "Contextual Online Learning for Multimedia Content Aggregation". *IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 17, NO. 4, APRIL 2015*

[2] S. Ren and M. van der Schaar, "Pricing and investment for online TV content platforms," *IEEE Trans. Multimedia, vol. 14, no. 6, pp. 1566–1578, Dec. 2012.*

[3] S. Song, H. Moustafa, and H. Afifi, "Advanced IPTV services personalization through context-aware content recommendation," *IEEE Trans. Multimedia, vol. 14, no. 6, pp. 1528–1537, Dec. 2012.*

[4] P. Kohli, M. Salek, and G. Stoddard, "A fast bandit algorithm for recommendations to users with heterogeneous tastes," *inProc. 27th Conf. AI, Jul. 2013, pp. 1135–1141.*

[5] S. D. Roy, T. Mei, W. Zeng, and S. Li, "Empowering cross-domain internet media with real-time topic learning from social streams," *in Proc. IEEE Int. Conf. Multimedia Expo, Jul. 2012, pp. 49–54.*

[6] S. Roy, T. Mei, W. Zeng, and S. Li, "Towards cross-domain learning for social video popularity prediction," *IEEE Trans. Multimedia, vol. 15, no. 6, pp. 1255–1267, Oct. 2013.*

[7] M. Naaman, "Social multimedia: Highlighting opportunities for search and mining of multimedia data in social media applications," *Multimedia Tools Appl., vol. 56, no. 1, pp. 9–34, 2012.*

[8] M. van der Schaar, J. Xu, and W. Zame, "Efficient online exchange via fiat money," *Econ. Theory, vol. 54, no. 2, pp. 211–248, 2013*

[9] L. Li, W. Chu, J. Langford, and R. E. Schapire, "A contextual-banditapproach to personalized news article recommendation," *inProc. 19th Int. Conf. World Wide Web, 2010, pp. 661–670.*

[10] M. Saxena, U. Sharan, and S. Fahmy, "Analyzing video services in web 2.0: A global perspective," *in Proc. 18th Int. Workshop Netw. Operating Syst. Support Digital Audio Video, 2008, pp. 39–44.*

[11] R. Mohan, J. R. Smith, and C.-S. Li, "Adapting multimedia internet content for universal access," *IEEE Trans. Multimedia, vol. 1, no. 1, pp. 104–114, Mar. 1999.*

[12] *A.*Slivkins, "Contextual bandits with similarity information," *inProc. 24th Annu. Conf. Learn. Theory, Jun. 2011, vol. 19, pp. 679–702.*

[13] G. Linden, B. Smith, and J. York, "Amazon.com recommendations: Item-to-item collaborative filtering," Internet Comput., *vol. 7, no. 1, pp. 76–80, 2003.*

[14] D. Bouneffouf, A. Bouzeghoub, and A. L. Gançarski, "Hybrid-greedy for mobile context-aware recommender system," *inAdvances in Knowledge Discovery and Data Mining. New York, NY, USA: Springer, 2012, pp. 468–479.*

[15] E. Hazan and N. Megiddo, "Online learning with prior knowledge," *inLearning Theory. New York, NY, USA: Springer, 2007, pp. 499–513.*