Journal of Zhejiang University SCIENCE
ISSN 1009-3095
http://www.zju.edu.cn/jzus
E-mail: jzus@zju.edu.cn

Enhancing the usage pattern mining performance with temporal segmentation of QPop Increment in digital libraries

CAO San-xing (曹三省)†[1,4], KLEIN R. Rody[2,3,4], LIU Jian-bo (刘剑波)[1]

(*[1]Information Engineering School, Communication University of China, Beijing 100024, China*)

(*[2]RCID, Zhejiang University, Hangzhou 310027, China*)

(*[3]SysCom Lab, University of Savoie, 73376 Le Bourget-du-Lac cedex, France*)

(*[4]Media Research, METIS Global Network, http://www.metis-global.org/*)

†E-mail: c3x@cuc.edu.cn; sanko_333@hotmail.com

Received Aug. 5, 2005; revision accepted Sept. 10, 2005

Abstract: The convergence of next-generation Networks and the emergence of new media systems have made media-rich digital libraries popular in application and research. The discovery of media content objects' usage patterns, where QPop Increment is the characteristic feature under study, is the basis of intelligent data migration scheduling, the very key issue for these systems to manage effectively the massive storage facilities in their backbones. In this paper, a clustering algorithm is established, on the basis of temporal segmentation of QPop Increment, so as to improve the mining performance. We employed the standard C-Means algorithm as the clustering kernel, and carried out the experimental mining process with segmented QPop Increases obtained in actual applications. The results indicated that the improved algorithm is more advantageous than the basic one in important indices such as the clustering cohesion. The experimental study in this paper is based on a Media Assets Library prototype developed for the use of the advertainment movie production project for Olympics 2008, under the support of both the Humanistic Olympics Study Center in Beijing, and China State Administration of Radio, Film and TV.

Key words: Media-rich, Digital library, Data migration, Media content, Log mining, QPop
doi:10.1631/jzus.2005.A1290 **Document code:** A **CLC number:** TP391

INTRODUCTION

With the development of Multimedia Data Pressure, Content-based Retrieval, Grid-based Information Processing, High-speed Internet and Massive Storage in recent years, media-rich digital libraries have become technically feasible and business-wise mature. The convenience in designing, implementing, deploying and upgrading of their applications is acting as the most important factor that drives content platforms practical. Application models of these portals can now be found in broadcasters' websites, online multimedia content providers and many Internet businesses (Song, 2001; Cao and Lu, 2001; Cao *et al*., 2003).

As indicated in Fig.1, an important issue that these systems are facing is the effective data migration model for the Hierarchical Storage schema of the media contents. Although HSM (Hierarchical Storage Management) and VSM (Virtual Storage Management) have respectively realized the multi-level model of data/content storage, and the consistency of storage access and application, they have also given birth to the problem of hierarchical data migration (Cao *et al*., 2004). In a massive storage system, the multiple storage modes/levels have necessitated the frequent migration of media data among them, according to the requirements of applications. Nevertheless, data migration is further emphasized by data warehousing, disastrous prevention backups, and heterogeneous integration.

In current industry, no data migration schemas with intelligent and effective scheduling are raised yet. As a result, frequent and random pushing and pulling of massive media data has been undermining the robustness and usability of the massive-storage-based systems in most Web application environments. Therefore it is essential that an intelligent data migration scheduling model be established on the basis of content objects' usage patterns, with the use of feature extraction and knowledge discovery, so as to ensure the effective functioning of massive media content portals on the Web.

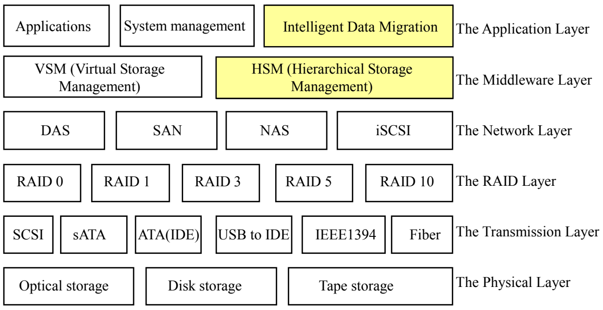


Fig.1 Technical framework of massive data storage (To see the relation of Intelligent Data Migration with neighbouring technologies)

RELATED WORK REVIEW

Data Migration has attracted the attention of many researchers in Information Processing and Computer Science since the last decade of the 20th century. In 1978, Todd (1978) of IBM posed the problem of data migration of geographically distributed databases, with the support of rights management, after which studies on different aspects of data migration were carried out. IEEE Storage System Standards Working Group published the model and infrastructure of massive storage in 1994, and after that, Data Migration studies are largely carried out in the environment of distributed massive storage, with consideration on network storage architectures, such as SAN, NAS and iSCSI. Current research on Data Migration is concentrated in 3 directions: study of data migration models based on engineering experiences; study of scheduling algorithms based on cybernetics, and study of system policies combining data migration with related technologies.

Khuller *et al*.(2004) established the polynomial-based temporal analysis model of data migration, and his implementation of the scheduling algorithm yielded the worst-case bound of 9.5. Gandhi (2004) established a 5.06 approximation algorithm for the Open Shop problem, which takes the complete migration time as the cost variable. This is much more advantageous than typical algorithms of 9.0 and 5.83.

Driven by the digitized and networked broadcasting media, and the convergent multimedia information services, the Content Management Platform has become an important part of the information industry's infrastructure. This emphasizes Hierarchical Data Migration as one of the key problems within the domain of multimedia information processing. Research on intelligent data migration in the integrated content service environment was presented in (Cao *et al*., 2004; Hu *et al*., 2005). And many case studies in application were carried out, for example Hu *et al*.(2004) has done a study focused on the automatic data migration schema based on TSM and DIVA of the IBM storage platform.

With the concrete progress of intelligence and cognitive sciences, data migration researches are introducing the Ontology-based Semantics and the Multi-Agent architecture, so as to form the Intelligent Distributed Data Migration Scheduling System, which is the part of the infrastructure of future Information Society.

INTELLIGENT HSM BASED ON QPOP INCREMENT CLUSTERING AND USAGE PATTERN DISCOVERY

Usage Pattern mining for HSM

Usage Pattern depicts how the users of a system access the resources or call the functional modules in a system, which is essential in both senses of HCI and AI. Within the massive media content environment under our consideration, Usage Patterns of media content objects stand for various possible ways of content access, such as querying, browsing, previewing and downloading, that happen frequently during the usage of the system.

As for HSM related problems in the massive media content environment, at least three levels of storage hierarchy should be considered, namely, Online Storage, Near-line Storage, and Offline Storage. These three storage levels are physically implemented as Disk Arrays, Data Tape Libraries and

CD-ROM Jukeboxes (Fig.2).

Disk Arrays can provide high speed access to stored media data, together with the seamless interfaces to content production systems such as NLE (Non-Linear Editing) and MAM (Media Assets Management). Therefore Disk Arrays are suitable for storing online media data, that is, the part of media data frequently accessed.

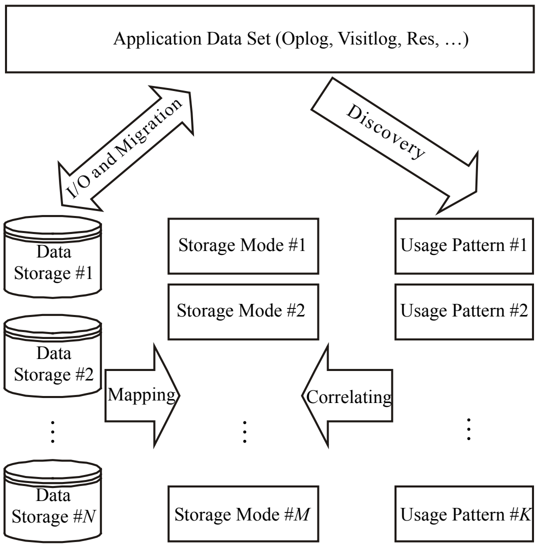


Fig.2 Intelligent HSM based on Usage Pattern Discovery

Data Tapes are used, in practice, to lower the storage mass' high total cost caused by the usually costly Disk Arrays. Although Data Tapes can only provide much inferior I/O speeds, they are economic and appropriate in practice, for storage of near-line media data, the part of media data that migrates into the online storage when necessary for access, and back into the Tapes after frequent usage, according to certain migration scheduling strategies.

For the storage of read-only and non-changeable media contents, such as the raw multimedia materials and archived productions, usually Compact Disk Libraries are used, which can also provide large storage volume with significantly lower price.

These different levels of storage have caused the problem of frequent duplication, removing and migration of large quantities of media data among the different storage levels. Consequently, considerable cost of system processing capability and resources has become not negligible, and the stability and robustness of the system can be evidently adversely affected. Therefore a hierarchical data migration model with high performance and effectiveness should be established, so as to assure and enhance the actual performance of the media content management platform. This requires the introduction of intelligent scheduling to the migration schema.

The data migration functioning system should not be a mechanical module that passively responds to migration commands caused by users' access requirements; it should be capable of smartly analyzing the usage patterns of the content objects, and thus generate a reasonable schedule for the content objects to be automatically migrated according to their possible usage, which can be predicted via existing usage pattern knowledge.

Either evident or implicit Usage Patterns can possibly be discovered by mining the preprocessed access log data. Migration scheduling with apriority and cluster considerations can thus be realized. This is to avoid or reduce the blindness and randomness of data migration, thus enhancing the quality of data migration functioning systems.

Usage pattern clustering based on QPop analysis

We use QPop (Query Popularity) as the feature for the mining of content objects' usage patterns. The QPops of content objects are the dimensionality-reduction results of the massive log data on various accesses to these content objects, usually in the form of dimension-limited vectors or scalars. During the use and running of the portal, a specific QPop Increasing Algorithm should be established according to the specific characteristics of the application environment. The Algorithm would quantifies the affects of all types of accesses on the content objects' user attention. Thus the increments of QPops can depict the dynamic characteristics of QPops and the content objects they stand for. The Intelligent Data Migration Scheduling System uses an effective KDD algorithm for clustering the QPop Increments of the large quantity of content objects, so as to find the usage patterns and then generate the input parameters for the migration engine.

Existing content platforms, most of which are Internet portals, usually define a simple QPop model (no matter what this model is called or shaped, for example Count of Visits, or Count of Favorites), and the accompanying QPop increase model (usually the

Increase-by-1 model), so as to find usage patterns (usually artificially or mechanically) of these content objects, and further plan and adjust the content architecture according to this knowledge.

In this paper we introduce standard KDD algorithms (with effective enhancements) into QPop Increment mining in media-rich information systems. For the convenience of studying the essential policy of Temporal Segmentation and Dimensionality Superinduction, we employed the basic QPop model (1-dimensioned, count of visits of the media content objects) and the basic QPop Increase model (count of visits increasing with seed=1). The experiment system was kept running for a working day in the actual application environment, when the QPop Increments are gathered for analysis.

C-MEANS-BASED QPOP INCREMENT CLUSTERING

As an important type of partitioning method for data clustering, C-Means is widely applied in data mining, and is usually chosen as the standardized algorithm for evaluating a mining strategy's effectiveness and comparing the performances of different mining strategies. In this paper, C-Means is employed as the kernel algorithm for clustering of QPop Increments, in order to find the usage patterns of content objects.

The C-Means-based QPop Increment Clustering method uses an algorithm kernel to cluster the QPop Increments of all the content objects during a specific time period. The clustering results depict the difference of usage characteristics of various content objects. For example, some content objects can be found to have a considerable increase in QPop during a period; while other objects hardly gain any QPop increase. C-Means is capable of providing these two groups of content objects with a satisfactory boundary. The migration strategies can thus be generated according to the clustering results. The workflow of this method is described below.

(1) For the given start and end time points t_0 and t, calculate the QPop Increments x_i=ΔQ_i=Q_{it}−Q_{it0} for each content object i at the time interval [t_0, t], and take {x_i} as the training dataset.

(2) For a definite cluster count k, take k values c_1, c_2, …, c_k from the training set {x_i} as the initial cluster centers.

(3) Put each sample x_i into one of the k clusters, according to the calculated Euclidean Distances indicated in Eq.(1)

$$\|x_i - c_l\| = \min_j \|x_i - c_j\| \tag{1}$$

(4) Adjust the cluster centers c_1, c_2, …, c_k, where c_i is calculated by:

$$c_i = \sum_{x_{l_i} \in C_i} x_{l_i} \Big/ N_i \tag{2}$$

and N_i is the cardinal number of the ith cluster.

(5) If the cluster centers c_1, c_2, …, c_k do not change anymore, the iteration ends; else go to Step (3).

For the comparison experiment carried out in this work, the QPop clustering algorithm with the above steps is firstly implemented to observe its performance difference with the enhanced approach described in the next Section.

IMPROVING QPOP INCREMENT CLUSTERING PERFORMANCE WITH TEMPORAL SEGMENTATION

The basic algorithm mentioned above simply processes the access log data during the studied interval into a 1-dimensioned integer, as the dimensionality-reduction result. Thus the changing tendencies of the media content objects' QPop values are ignored. Blind to the changing tendencies, the QPop Increment values will not be competent for depicting the usage characteristics of the content objects, especially when our study background is the media-rich content environments, where the features of the content objects are extremely complicated. Therefore this paper introduces an improved approach of QPop Increment modelling and then the related clustering strategies, so as to enhance the performance of QPop Increment clustering. The essential part of the improved QPop Increment model is the dimensionality-superinduction via QPop segmentation in the time domain.

We locate additionally m interval points between the start time t_0 and the end time t, so as to segment the studied time interval (and the QPop Increment on it) into m+1 partial increment values. Thus the QPop Increment of each content object at the studied time interval is reshaped into an (m+1)-dimensioned vector. Our approach then carries out the iterative clustering process with the kernel algorithm (in this paper C-Means). The advantage of temporal-segmented multi-dimensioned QPop Increment vectors is that they take the specific increasing tendencies of the QPop values in the given time interval. The resolution of the increasing tendencies to be visible can be adjusted by changing the interval points count m, where a larger m would result in a more detailed view of the QPop increasing tendencies while increasing the computational complexity. The capability of considering specific increasing tendency of the content object's QPop value is therefore to act as a good factor in the usage pattern mining process, which would be indicated in the experiment results in the next Section.

The specific process of the enhanced approach is described below.

(1) For the given start and end time points t_0 and t, and the given interval points count m, calculate the time segmentation interval τ, where $\tau=(t-t_0)/(m+1)$. Thus we get the time value of the interval points $t_j=t_0+j\tau$, where j=1, 2, …, m+1.

(2) Calculate the segmented QPop Increments $x_{ij}=\Delta Q_{ij}=Q_{it_j}-Q_{it0}$ for each content object i on the time interval $[t_0, t]$.

(3) Take the set of (m+1)-dimensioned vectors $\boldsymbol{x}_i=[x_{i1}, x_{i2}, \ldots, x_{i(m+1)}]^{\mathrm{T}}$ for all is as the training dataset.

(4) For the cluster count k, take k (m+1)-dimensioned vectors $\boldsymbol{c}_1$, $\boldsymbol{c}_2$, …, $\boldsymbol{c}_k$ from the training set $\{\boldsymbol{x}_i\}$ as the initial cluster centers.

(5) Put each sample $\boldsymbol{x}_i$ into one of the k clusters with the cluster center $\boldsymbol{c}_i$, according to the calculated Euclidean Distances indicated in Eq.(3)

$$\left\|\boldsymbol{x}_i-\boldsymbol{c}_l\right\|=\min_j\left\|\boldsymbol{x}_i-\boldsymbol{c}_j\right\| \tag{3}$$

(6) Tune the cluster centers $\boldsymbol{c}_1$, $\boldsymbol{c}_2$, …, $\boldsymbol{c}_k$, where $\boldsymbol{c}_i$ is calculated by:

$$c_{ip}=\sum_{x_{l_i}\in C_i} x_{l_i p}\Big/N_i \tag{4}$$

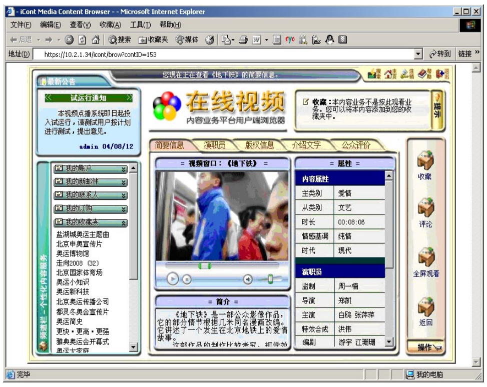
and N_i is the cardinal number of the ith cluster, and c_{ip} is the pth component of $\boldsymbol{c}_i$.

(7) If the cluster centers $\boldsymbol{c}_1$, $\boldsymbol{c}_2$, …, $\boldsymbol{c}_k$ are stabilized, the iterative process ends; else go to Step (4).

We have implemented the above approach into the experimental system, and carried out the comparison experiment on QPop Increment mining.

EXPERIMENT RESULTS AND ANALYSIS

Taking the standard C-Means as the kernel algorithm, while using the basic QPop increase model of "Visit Count Increase by 1", we established an experimental system of media content objects' usage pattern discovery (Fig.3). Inside the system, two sets of QPop Increment mining algorithms are implemented, namely, the basic algorithm, and the Temporal-Segmentation-based improved algorithm.



https://10.2.1.34/icont/brow?contID=153

您现在正在查看《地下铁》的简要信息。

试运行通知

本视频点播系统即日起投入试运行，请测试用户按计划进行测试，提出意见。

admin 04/08/12

收藏：本内容业务不是按此观看业务。您可以将本内容添加到您的收藏夹中。

盐湖城奥运主题曲
北京申奥宣传片
奥运博物馆
走向2008 (32)
北京国家体育场
奥运小知识
奥运新科技
北京奥运传播公司
都灵冬奥会宣传片
奥运简史
更快・更高・更强
雅典奥运会开幕式
奥运大家庭

= 视频窗口：《地下铁》 =

= 简介 =

《地下铁》是一部公众影像作品，它的部分情节根据几米同名漫画改编。它讲述了一个发生在北京地铁上的爱情故事。
这部作品的制作比较专业，…

= 属性 =

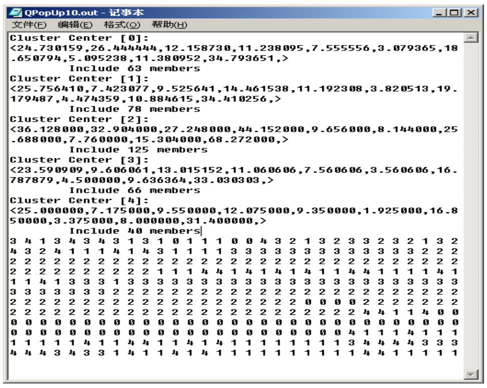
内容属性	
主类别	爱情
从类别	文艺
时长	00:08:06
情感基调	纯情
时代	现代
演职员	
监制	周一楠
导演	郑凯
主演	白鸥 张萍萍
特效合成	洪伟
编剧	游宇 江珊珊

Fig.3 The UI of the experimental content portal

For the specific experiment process, we gathered an expert team made up of 12 professionals in A/V multimedia contents production, so as to simulate the networked digital production team in actuality. The team was assigned the task of searching, browsing and accessing the experimental content library in the natural way as if a real collaborative production project is being carried out. The content library contains 372 media content objects of various types including video clips, audio clips, movies and TV productions. The experiment system records the detailed access log and produces the QPop Increments for the two mining algorithms. Both algorithms were run on the QPop Increments produced by the same experimental process, and we compared the difference of the mining results of the two approaches. The main parameters of the experimental environment and process are: Cardinal number of the content object set: 372; Primary content category space: {0102, 020201, 020202, 030101, 030102, 030103}; Increasing parameter of the basic QPop Increment Model: 1; Start and end time of the experiment: 9:30~15:30; Highest temporal segmentation dimension: 12; Kernel clustering algorithm: C-Means; Cluster centers: Self-random Initial Values; Destination pattern (cluster center) count: 5; Number of clients: 12.

The Destination Pattern (Cluster Center) Count is set to 5, which is for the purpose of simulating the usual 5 storage modes in actual Media Assets (Content) Management environments, namely, the Data Tape Library, the CD Jukebox, the SAN, the NLE-based SCSI Disk and the NLE-based IDE Disk.

The output data format of the clustering algorithm is indicated in Fig.4. The upper part of the output data is the multi-dimension coordinate of the cluster centers ($\boldsymbol{c}_i$), together with the count of objects that each cluster owns. The lower part of the data indicates the cluster IDs each content object belongs to (totally 372 values).



```
Cluster Center [0]:
<24.730159,26.444444,12.158730,11.238095,7.555556,3.079365,18
.650794,5.095238,11.380952,34.793651,>
        Include 63 members
Cluster Center [1]:
<25.756410,7.423077,9.525641,14.461538,11.192308,3.820513,19.
179487,4.474359,10.884615,34.410256,>
        Include 78 members
Cluster Center [2]:
<36.128000,32.904000,27.248000,44.152000,9.656000,8.144000,25
.688000,7.760000,15.304000,68.272000,>
        Include 125 members
Cluster Center [3]:
<23.590909,9.606061,13.015152,11.060606,7.560606,3.560606,16.
787879,4.500000,9.636364,33.030303,>
        Include 66 members
Cluster Center [4]:
<25.000000,7.175000,9.550000,12.075000,9.350000,1.925000,16.8
50000,3.375000,8.000000,31.400000,>
        Include 40 members
3 4 1 3 4 3 4 3 1 3 1 0 1 1 1 0 0 4 3 2 1 3 2 3 3 2 3 2 1 3 2
4 3 2 4 1 1 1 4 1 4 3 1 1 1 1 3 3 3 3 3 3 3 3 3 3 3 3 3 2 2 2
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
2 2 2 2 2 2 2 2 2 2 1 1 1 4 4 1 4 1 4 1 4 1 1 4 4 1 1 1 1 4 1
1 1 4 1 3 3 3 1 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
3 3 3 3 3 3 3 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 0 0 0 0 2 2 2 2 2 2 2
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 4 4 1 1 4 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 4 1 1 1 4 1 1 1
1 1 1 1 1 4 1 1 4 4 1 1 4 1 4 1 1 1 1 1 1 1 1 3 4 4 4 4 3 3 3
4 4 4 3 4 3 3 1 4 1 1 4 1 4 1 1 1 1 1 1 1 1 1 1 4 4 1 1 1 1 1
```

Fig.4 Format of the usage pattern clustering output

The clustering experiments were carried out under the multiple situations of the basic approach, and the improved approach with m=1, 2, …, 11, where the dimension of the QPop Increment was respectively increased to 2, 3, …, 12. Altogether there were 12 groups of results. We analyzed and compared the essential clustering performance indication, the Cluster Cohesion, which is quantified by calculation of Eq.(5):

$$\Psi=\sum_{j=1}^{k}\frac{\sum_{x\in C_j}\left\|\boldsymbol{x}-\boldsymbol{c}_j\right\|}{N_j} \tag{5}$$

where k is the destination pattern (cluster) count (in this experiment, k=5); N_j is the cardinal number of the jth cluster; $\|\boldsymbol{x}-\boldsymbol{c}_j\|$ is the Euclidean distance of the member content object's QPop Increment vector $\boldsymbol{x}$ and the center $\boldsymbol{c}_j$ of the cluster C_j this object belongs to. Listed in Table 1 are Ψs of each group of results.

Table 1 Cluster cohesions quantified with Ψ

Dimension	Ψ
1 (Basic approach)	21.908801
2	18.922970
3	17.135127
4	14.133139
5	11.126135
6	10.689538
7	9.405826
8	9.186779
9	8.331318
10	7.490788
11	7.355276
12	7.278397

Fig.5's visualization of the Ψs in a graph where the horizontal axis stands for dimension and the vertical one for the value of Ψ indicates that the improved approach can yield considerably better cluster cohesion. The Ψ value decreases faster when the superinduction dimension increases from 2 to 5. For the case where the dimension is 5, the Ψ value is only 50.783860% of that in the basic approach case. The Ψ value decreases continuously when the dimension increases from 5 to 12. When the dimension reaches 12, the Ψ value is only 33.221338% of that in the basic case. Better cluster cohesion depicts more clearly the cluster boundary, and simultaneously, better aggregation of content objects to the clustered Usage Patterns. This would further result in more concise configuration and better implementation of the data migration policies, which would be advantageous to the performance of the Web-based content library.

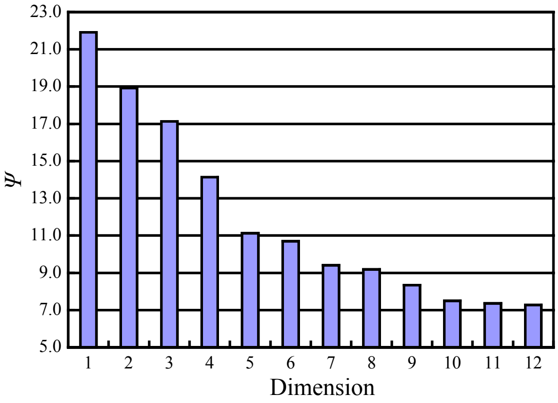


Fig.5 Visualized Comparison of Ψs

The experiment results also indicated that, when the dimension increase is relatively high, the Ψ value tends to change less. For the experiment, the clustering algorithms, both the basic one and the improved one, can finish the clustering process within several hundred ms or several seconds. This indicates a trade-off between clustering quality and computational complexity. The empirical strategy for the optimization of this trade-off is to chose a dimension of 5~6, where the Ψ value decreases relatively fast but tends to become much slower when the dimension still goes up.

CONCLUSION AND FUTURE WORK

In this work we carried out an experimental study on the essential technology of intelligent data mining in media-rich digital libraries. These types of information systems have a promising future application, especially during the convergence of media, communication and computing networks, when the massive digital contents from media organizations are to be distributed via the content-based and semantics-oriented Web environment.

Specifically, this paper proposed an improved approach in mining the QPop Increment for content object usage patterns. This approach segments the QPop Increment during a given time interval into multiple components of a vectorized multi-dimension value. We designed and implemented an experimental portal running both the basic C-Means-based QPop Increment clustering algorithm and the improved vector clustering algorithm. The multiple groups of the experiment results are analyzed and compared. It is verified that the improved algorithm can enhance the clustering performance to result in clearer usage patterns advantageous for basing data mining policies on.

In work following this project, we will detailedly study the multi-dimensioned QPop models and the multiple-factor-driven QPop increase models. The modelling of media content semantics and ontology-based feature description is related with the study on the better QPop model and QPop increase model. For the mining procedure, fuzzy clustering algorithms, among which are FCM, PIM, FCS and RCP, will be used, in order to reduce the inappropriate object migration during the iterative clustering process by introducing the fuzzy memberships of objects to clusters. In aspect of application, the following study is aimed at providing a more effective and operable intelligent data migration engine for the Media Assets/Content Management environments.

References

Cao, S., Lu, R., 2001. Large-Scale TV Station OA Application Systems Based on Intranet and Web. Proc. CIEYC2001, Beijing Broadcasting Institute Press, Beijing, p.234-237.

Cao, S., Xu, J., Gao, F., 2003. Media Enterprise Resource Planning: Concept and Application Framework. Proc. 8th Intl. Symp. on Broadcast Tech., Hong Kong, p.120-124.

Cao, S., Klein, R., Zheng, G., Geng, W., 2004. Managing Uncertainty in Media Content Platforms. Proc. 4th Intl. Symp. on Management of Technologies. Zhejiang University Press, Hangzhou, p.171-175.

Gandhi, R., 2004. Improved results for data migration and open shop scheduling, *LNCS*, **3142**:658-669.

Hu, L., Meng, F., Hu, M., 2004. A Dynamic Load Balancing System Based on Data Migration. 8th International Conference on Computer Supported Cooperative Work in Design (IEEE Cat. No.04EX709), **1**:493-499.

Hu, W., Cao, S., Li, D., 2005. An improved algorithm of media content usage pattern mining based on temporal segmentation and dimensionality superinduction of QPop increases. *Computer Science*, **22**(7B):178-180, 208.

Khuller, S., Kim, Y.A., Wan, Y.C., 2004. Algorithms for data migration with cloning. *SIAM Journal on Computing*, **33**(2):448-461.

Song, Y., 2001. Technology and Choices of Media Assets Management. Proc. 8th Annual Academic Conference of CSMPTE, Beijing, p.4-20.

Todd, S., 1978. Automatic data migration in distributed database system. *IBM Technical Disclosure Bulletin*, p.387-388.