PERSONALIZED EVENT-BASED NEWS VIDEO RETRIEVAL WITH DYNAMIC USER-LOG

Ming Li[1, 2, 3] *Yantao Zheng*[1] *Shi-Yong Neo*[1]
Xiangdong Wang[2] *Sheng Tang*[2] *Shou-Xun Lin*[2]

[1] School of Computing, NUS
[2]Key Laboratory of Intelligent Information Processing, ICT, CAS
[3]Graduate University of Chinese Academy of Sciences

ABSTRACT

Personalization especially in the domain of information retrieval is essentially important, as users might pose the same query even when they are searching for different information. It is thus necessary to create a retrieval engine which takes into consideration the dynamic information needs of different users. This paper presents our personalized news video retrieval engine, which exploits the individual user's previous browsing history to customize and enhance their future search results. Specifically, the system utilizes the news topic hierarchy, a hierarchical news topic structure derived from unsupervised clustering on the news video corpus and event entities from news video and online news articles. We then dynamically project user's browsing history onto this topic hierarchy to provide the basis for re-ranking relevant news videos. This system is tested on one month of TRECVID 2006 dataset consisting of 80 hours news video and found to return results in a more intuitive and personalized manner.

Index Terms— Personalized retrieval, user-log

1. INTRODUCTION

The explosion of multimedia content has generated new requirement for more effective multimedia retrieval. The same queries post from different users may actually target different information needs. It is thus necessary to personalize search for different users so that the returned results can be more satisfactory to fulfill searchers' need. Ghosh et al. [1] provided personalized ranking of results in video retrieval using implicit user feedback from click-through. A Bayesian network was trained from the click-through data to model personalization for re-ranking retrieval results. However, the association knowledge which Bayesian network is trying to model may not be stable for the semantic gap effect the relationship between concept and video features. Zhang et al. [2] applied the personalized retrieval on sports video, in order to acquire user's preference. The relevance feedback is applied and semantic video annotation is prerequisite. There are also other traditional methods of obtaining user's preference by asking users to manually enter their choices but they are not effective in modeling the changing needs of users. The challenge here is how to effectively model the dynamic needs automatically.

In this work, we attempt to utilize browsing history of users of news video in a temporal fashion to refine future search results. The intuition is to gather the dynamic information needs through the news videos which the users have seen, as they might be interested in similar events or topics. We propose a temporal multi-stage clustering that performs unsupervised clustering on the news materials. The resultant structure from this clustering is a Topic Hierarchy that is organized based on news event. The user browsing history is then projected onto the Topic Hierarchy to obtain the searching trend on topics and events. To model the dynamic changing needs, this personalization score is biased towards events that are last seen by the user. Experiments on TRECVID 2006 dataset showed that our personalized news video search engine gave a significant improvement in retrieval experience when compare to general search engine.

2. TEMPORAL MULTI-STAGE CLUSTERING

In this Chapter, we will discuss our multi-stage event clustering which uses event entities from news materials in a hierarchical *k*-means clustering. The primary source of obtaining event entities is through the use of speech transcripts which are made available from automated speech recognizer (ASR). However, due to the erroneous nature of speech transcripts, it is unlikely to obtain full aspects of an event. This limitation prompts the use of relevant external resources, in particular, the parallel text online news resources to supplement ASR text of news video. We propose to leverage a combination of news articles and news video stories in the same clustering space. The rationale is to make use of the innate associations between event entities from both sources of news.

Performing a single clustering on the entire news corpus [3] is straightforward and simple. However, such clustering process has two major drawbacks: lacking of robustness to outliers and computationally expensive. In order to tackle the two issues above, it is necessary to partition the data into suitable sizes for clustering. We therefore create temporal partitions so that the clustering can be performed on a smaller scale. The drawback from partitioning is that the overall structure becomes disconnected. We then further make use of threading across the events to re-connect these partitions.

2.1. Leveraging External News Articles

The use of news articles is important in bridging missing semantic entities of speech transcripts. For example, considering the following two news videos (story1 and story2) of a same event but having missing entities due to speech transcription errors or machine translation. Text of story1: "*Chemical factory explosion kills more than ten people in California*"; Text of story2: "*Blast in Western United States ... refinery death toll to 14*". The text of story2 did not mention California and have many terms which are different to story1. In addition, we found that story2 actually have a higher similarity value to story3, which have terms like "*Death toll of the tsunami in Western Java rose to more than thousands, the red cross and the United Nations are ....*" reported in the same period. In this case, it is possible that story1 and story2 may be clustered wrongly into different clusters. It is important to provide a semantic bridge so that story1 and story2 can be clustered together.

We leverage news article of the "explosion event" to provide better and more complete description like text from article: "*The number of death in the oil refining chemical factory located Western United States, California rose to more than 10. The blast was believed to be caused by ...*" We see that text from this article overlaps with both story1 and story2. When applied in clustering, this causes the cluster centre of this "explosion event" to be shifted in a dimension closer to both story1 and story2, thus creating a high probability of having story1 and story2 in the same cluster.

2.2. Multi-stage Event Clustering Framework

By using the story boundaries from [4] to segment the news video, we perform hierarchical *k*-means clustering on corpus containing both news video and news articles (see Figure 1). The intuition for multi-stage clustering is to use only story context for first stage clustering and a combination of context and visual for second stage clustering. Clustering at the first stage uses only text entities from both news articles and news video stories in a hierarchical k-means clustering framework. After obtaining the initial clusters, visual features, namely high level features from news video stories are then added at the second stage for further clustering of news video stories.

Text Similarity: The distance measurement employed during clustering is the cosine similarity used in vector space text retrieval model [5], which is commonly used in measuring similarity in text mining.

Eqn 1 shows the cosine similarity formula:

$$Text\_Sim(v_i, v_j) = \frac{v_i . v_j}{|v_i| . |v_j|} \qquad (1)$$

where $v_d = [w_{1,d}, w_{2,d}, ..., w_{t,d}]^T$, $w_{t,d} = tf.idf$. The representative vector v of the news materials is the list of event entities extracted from the news articles and news video obtained from Name entity extractor in [6].

Visual Similarity. Even though most event entities can only be obtained from speech, there are many audio and visual elements in the news video that can be important. For example: audio signatures like "engine noise" can indicate an aircraft taking off, "clapping or cheering" can indicate a large crowd; or visual scenes of "fire" can indicate events like fire outbreak or forest fire. This important information may not be available from text. We make use of the set of 50 high level video features in [7] for further representative features of the news video.

$$HLF\_Sim(v_i, v_j) = Norm(Euc\_HLF(v_i, v_j)) \qquad (2)$$

Euc_HLF() computes the Euclidean distance between two news video stories using their 50-dimension high level feature. These values are normalized before they are used in the clustering process.

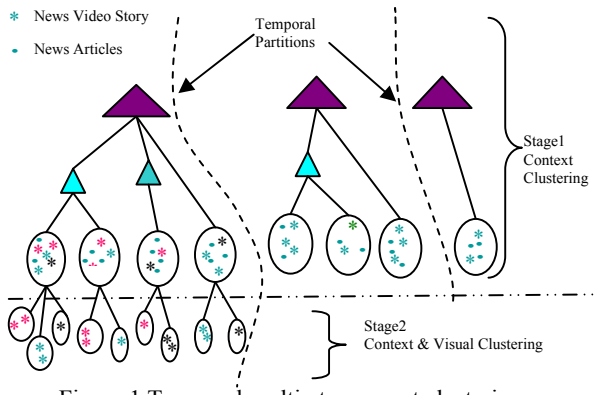

Figure 1 Temporal multi-stage event clustering

2.3. Temporal Partitioning and Threading

After obtaining the various hierarchical clusters for each temporal partition using technique from previous Section, we need to link and thread these hierarchical structures so that they can be used as a single structure during retrieval. The partition on news video corpus is based on temporal information (date) from the video. This is because events are usually time dependent and news relating to the same event tend to be concentrated in the same period of time. To preserve the context, we further introduce an overlap between the partitions so that locality information can later be induced. In our study, we found the temporal partitions of five days with two days overlap working well. To combine the disconnected structure, we leverage the event entities in the news video in the cluster [8]. 3 types of links are considered (see Figure 2): "identical", "near-duplicate" and "high-similarity".

The first type, "identical", occurs when two clusters contains the same set of news video stories. This is possible due to the overlapping partition. In this case, a direct link will be created between the 2 clusters. The second type, "near-duplicate", occurs when Sim() score > δ_n. In this case, the two clusters must have overlapping news video instances and high similarity. For this, we will create a fusion link for this type of relation. The third type, "high-similarity", occurs when clusters have a similarity value of above preset threshold δ_s.

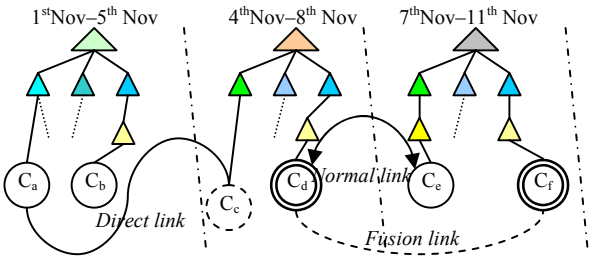

Figure 2 Threading clusters across temporal partitions

3. PERSONALIZATION AND RETRIEVAL

The personalization step employs the user's video browsing history to build the P-TH (personalized-Topic Hierarchy). The P-TH will act as a personalization mechanism to re-adjust news videos to be returned to the individual user from automated retrieval. To appropriately model the changing needs of each user, the P-TH uses only the last *n* browsed news video. In addition, extra weights are given to cluster or nodes of the P-TH which contains the last seen news videos.

3.1. Modeling Browsing History Graphically

The P-TH follows the exactly the same graphical structure resultant from the Topic Hierarchy except that the nodes contain numerical values as seen in Figure 3.

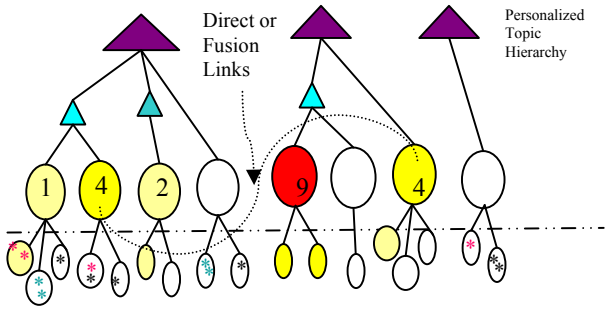

Figure 3 P-TH for user-A

At the very beginning, the value of each cluster/node is preset to zero. Once the user browses a news video, the parent node which contains the news video is added with some scores. In our implementation, the value of each node is determined by the last *n* viewed news videos by the user. In addition, higher weights are given to nodes which inherit the last seen videos rather than those which are viewed much earlier. According to this rational, the value representative of the nodes will indicate the individual interest level of user to that particular kind of news event or topic. These values can be used as the personalized retrieval values, which are fused together for retrieval in Section 3.2.

We leverage on the links in the topic hierarchy: nodes that can be reached by a single direct or fusion link are also updated since they might be topics or events, in which the user is interested. In Figure 3, the nodes containing the numerical value 4 is one such example. The algorithm for updating the P-TH is shown in Figure 4. Let the list of last seen videos for user-A be L_A and v_i is a video in L_A. In our implementation, we only consider the last 30 news video seen by the user. This number is arbitrary chosen but found to work well through the experiments.

1. Initialize P-TH for user-A,
   Set all node to 0
   Set score = *k*
2. For last *n* browsed videos {
   Add *score* to parent node containing v_i
   Update score of nodes by fusion or direct links
   score= score--
   }
3. Normalize nodes values [0..1]

Figure 4 Algorithm for creating P-TH

3.2. Incorporating Personalization into Retrieval

As an essential part of video search, a reliable automatic search engine should supply a high-quality initial rank list for the retrieval. We employ our automatic search techniques based on our previous work [8] to perform the initial retrieval. We induce and extract query-information like query terms, query-class and query-HLF (high-level feature) from the text query supplied by users. We incorporate the initial retrieval score *auto()* and normalized nodes value in P-TH *per()* to get a the final personalized retrieval S*core()*.

$$Score(v) = \alpha \cdot auto(v) + (1-\alpha)\, per(v) \qquad (3)$$

where per() directly uses the value of the parent node containing video v. The parameter $\alpha \in [0,1]$ can be flexibly defined by the user so as to bias towards relevancy or user searching preference.

4. EXPERIMENTS AND RESULTS

To test the effectiveness of our system, we make use of the TRECVID 2006 news video dataset. We craft a subset consisting of 80 hours news video collected in Nov 2005. Two sets of experiments are designed. The first set of experiment investigates the clustering performance of the technique proposed in this paper. The second experiment demonstrates a retrieval test with 10 users on personalize retrieval.

4.1 Experimental Results of Clustering

With the various techniques introduced above, it is important to understand how they can individually affect clustering performance. For evaluating the clustering performance, we manually screen through the video and perform topic-based grouping. This is done using approximately 10 hours of news video. The experiments are designed to test the effectiveness of: (1) usage of parallel news articles for clustering; and (2) usage of temporal partitions.

The quality of clustering is determined by analyzing the entire hierarchical structure. This is often done by using a measure that takes into account the overall set of clusters that

are represented in the hierarchical tree. One such measure is the F_{Score} measure, employed by [9, 10] which is modified from the usual F1 measure. We follow the evaluation as in [9]. The first series of runs are constructed as follows and the results are tabulated in Table 1

baseline: (without parallel news and temporal partitions)
T) baseline + temporal partitions
P) baseline + parallel news
TP) T + parallel news
TPH) TP + high level features

Table 1 Performance of clustering (percentage of improvement over baseline)

T2006	Baseline	T	P	TP	TPH
F_{Score}	0.298	0.388	0.345	0.455	0.545

From Table 1, we can draw the following observations. First, the use of parallel news is effective as can be seen in improvements in clustering performance for P over baseline. Second, significant improvement can be seen from the addition of temporal partitions that can be seen from run T and TP. This is mainly attributed to the nature of news video that is time dependent in nature. A significant difference in time usually means distinctive events. The run which uses high level features yield the best performance, demonstrating that the multi-stage clustering is effective.

4.2 Experiments on Personalize Retrieval

This series of test investigate the effectiveness of the personalize retrieval aspect of the system Ten users who have prior knowledge in online news video retrieval are selected for the experimentation. The users will be asked to carry out retrieval on the following three systems to determine their effectiveness.

U) General automated retrieval based on [8]
UA) U with (Eqn 3) without fusion and direct links
US) U with (Eqn 3) with fusion and direct links

The testing corpus is divided into two temporal portions, first two weeks and next two weeks of news video. The system will pick up the personalization values when the user search on the first portion and subsequently apply it on the second portion. This is rationale as it follows exactly the methodology of using the browsing history for future queries. Note that personalization scores will not be applied for the first portion of retrieval and the users are **not** told of the difference between the various systems specifications or whether personalization is used. Thereafter, the users are asked to rank the system in terms of retrieval accuracy from a scale 1 to 5 (5 as having best performance).

From Table 2, we can see that the UA and US runs yield better scores than U run for the 2nd portion runs. In particular, US runs have significantly better results over U runs. Some user feedback that the system is capable of understanding their needs like giving news on soccer instead of general sports news when there are searching for sports news. This effectively means that the personalization methodology works well for searcher. In addition, the user agreement based on the first 2 weeks of video shows that the experiment is carried in an un-bias fashion.

Table 2 Mean user rating for personalize retrieval

Video	U	UA	US
First 2 wks (mean)	3.2	3.1	3.3
Next 2 wks (mean)	3.5	4.0	4.3

5. CONCLUSION

A temporal multi-stage clustering is introduced together with the use of news event topic hierarchy for effective personalized news retrieval. Our proposed framework enables us to obtain searching trends which effectively enhances personalized retrieval performance. Experiments on TRECVID 2006 dataset showed that our personalized news video search engine produces statistically significant improvements on modeling users' preferences over automated video search engines.

6. ACKNOWLEDGEMENTS

We would like to thank Professor Tat-Seng Chua for his excellent suggestions and discussions. This work was supported in part by the National Basic Research Program of China (973 Program, 2007CB311100), the National High Technology and Research Development Program of China (863 Program, 2007AA01Z416), the National Nature Science Foundation of China (60773056), the Beijing New Star Project on Science & Technology (2007B071).

7. REFERENCES

[1] Hiranmay Ghosh et al., "Learning ontology for personalized video retrieval" in *International Multimedia Conference 2007.*

[2] Y. Zhang et al."Personalized Retrieval of Sports Video" in *MIR'07, September 28–29, 2007, Augsburg, Bavaria, Germany.*

[3] S-Y. Neo, Y. Zheng, T-S. Chua, Q. Tian, "News Video Search with Fuzzy Event Clustering using High-level Features" *ACM Multimedia 2006, Santa Barbara, USA, 23-27 Oct 2006.*

[4] W.H. Hsu, et al"Columbia-IBM News Video Story Segmentation in TRECVID 2004", *Columbia ADVENT Technical Report, New York 2005*

[5] G. Salton, A. Wong, and C. S. Yang (1975), "A Vector Space Model for Automatic Indexing," *Communications of the ACM, vol. 18, nr. 11, pages 613–620.*

[6] H. Yang, L. Chaisorn, Y. Zhao, S.-Y. Neo, and T.-S. Chua. "VideoQA: question answering on news video." In *Proc. of the 11th ACM MM, pages 632–641, 2003.*

[7] Tat-Seng Chua, Shi-Yong Neo et al. "TRECVID 2006 by NUS-I2R" In *TRECVID 2006, NIST, Gaithersburg, Maryland, USA, 13-14 Nov 2006.*

[8] S.-Y. Neo, Y. Ran, H.-K. Goh, Y. Zheng, T.-S. Chua, J. Li, "The Use of Topic Evolution to help Users Browse and Find Answers in News Video Corpus," *ACM MM 2007, Augsburg, Germany, 23-29 Sep 2007.*

[9] B. Larsen and C. Aone. Fast and effective text mining using linear-time document clustering. In *Proc. of the Fifth ACM SIGKDD Int'l Conference on Knowledge Discovery and Data Mining, pages 16–22, 1999.*

[10]Y. Zhao, G. Karypis, "Evaluation of hierarchical clustering algorithms for document datasets" *Conf Information and Knowledge Management, pp515-524,McLean, Virginia, USA, 2002*