Feature Subspace Selection for Efficient Video Retrieval

Anuj Goyal, Reede Ren, and Joemon M. Jose

Dept. of Computing Science, University of Glasgow, Glasgow, G12 8RZ, UK
{anuj,reede,jj}@dcs.gla.ac.uk

Abstract. The curse of dimensionality is a major issue in video indexing. Extremely high dimensional feature space seriously degrades the efficiency and the effectiveness of video retrieval. In this paper, we exploit the characteristics of document relevance and propose a statistical approach to learn an effective sub feature space from a multimedia document collection. This involves four steps: (1) density based feature term extraction, (2) factor analysis, (3) bi-clustering and (4) communality based component selection. Discrete feature terms are a set of feature clusters which smooth feature distribution in order to enhance the discrimination power; factor analysis tries to depict correlation between different feature dimensions in a loading matrix; bi-clustering groups both components and factors in the factor loading matrix and selects feature components from each bi-cluster according to the communality. We have conducted extensive comparative video retrieval experiments on the TRECVid 2006 collection. Significant performance improvements are shown over the baseline, PCA based K-mean clustering.

1 Introduction

Video retrieval attracts great interest from both industry and academic fields. However, existing retrieval systems have a high computational complexity. This is due to two reasons. First, a video document consists of many media modalities such as audio track, textual tags and visual frames. Second, video contents and associated semantics have no direct correlation with low-level features. Moreover, Wang [10] asserts that retrieval is not a simple discrimination on local features but a measurement of uncertainties among possible relevant documents. This indicates that the noise in the feature space will result in extra complexity in the measurement of document relevance and degrade retrieval performance. The optimization on video document representation is therefore essential to improve the effectiveness as well as the efficiency of a video retrieval system.

In this paper, we exploit techniques from statistical information retrieval to learn an efficient feature subspace from media collections. These techniques are: (1) dimension based density normalization; (2) factor analysis in the normalized feature space; and (3) bi-clustering for subspace allocation. This is because the theory of information retrieval has already developed many hypothesis on the feature distribution. These knowledge has not been used by traditional dimensionality reduction methods such as principle component analysis (PCA).

S. Boll et al. (Eds.): MMM 2010, LNCS 5916, pp. 725–730, 2010.

Moreover, retrieval is a statistical decision based on the difference of feature distributions in both document collection and queries [11]. Statistical information from document collections may facilitate the creation of a better feature space. We hence start our work from density normalization by projecting continuous distributed features to a set of discrete variables called feature terms. This projection will maximize the discrimination between documents in a collection. We then use factor analysis to compute the correlation between dimensions in the feature term space and get the loading matrix. To make groups in loading matrix we propose to apply bi-clustering on it. From each bi-cluster we select the component which has minimum communality as a feature subspace for document representation and for relevance computation.

The remainder of this paper is organized as follows. Section 2 brings a brief overview about the literature of textual term and feature subspace selection in content-based video retrieval. Our approach for feature subspace selection is presented in Section 3. Experiment configuration and evaluation results are stated in Section 4. Discussion and conclusion are found in Section 5.

2 Related Work

In this paper, we try to exploit techniques in statistical information retrieval for feature subspace selection. Many essential issues require explanation, such as document representation and relevance estimation.

2.1 Term Distribution

As an important part of term weighting, text term distribution has been well addressed to justify text retrieval models [3,1]. Many hypothesis have been proposed to simulate a general term distribution. Harter *et al.* [3] declare that a term should follow a 2-Poisson distribution, since term appearance is a Boolean random phenomena with a low average arrival rate. This model is extended by Margulis *et al.* [8] who test N-Poisson distributions. The authors hypothesize that N-Poisson might have provided a more precise estimation than a 2-Poisson hypothesis, if a term actually followed a Poisson-like distribution. Several class numbers from two to seven are evaluated on real document collections, although no optimized solution is reached. Amati *et al.* argue that the joint probability of multiple terms is so small that a simple uniform distribution is good enough for the term distribution modeling.

In multimedia retrieval, several approaches have been proposed to extract term-like media features, such as high-level features and SIFT-based local features. Although the distribution of these term-like media features has not been well studied, it is interesting to exploit effective textual models for multimedia documents.

2.2 Feature Subset Selection

Feature subset selection (FSS) is an important optimization approach for multimedia retrieval [4,2]. This technique aims to select the most effective feature components in the document representation without losing performance. The key operation of FSS is to estimate the discriminative power of a feature component. A multi-layer perceptron network is proposed in [6] to classify variables into two groups, effective and ineffective, where a stepwise discrimination is used as input. Principal Component Analysis (PCA) is also widely used to find an optimal solution in data representation. As will be shown later, these methods have many disadvantages.

In this paper, we use factor analysis (FA) for un-supervised selection of feature term components which overcomes the shortcomings of PCA. Furthermore, we apply bi-clustering on a loading matrix to group different feature components. Bi-clustering can separate factor subgroups more efficiently, because unlike K-means clustering, it is flexible to choose any dimensions of a feature as well as any combination of factors, from the loading matrix.

3 Methodology

In this section, we describe our method for feature subspace selection, including density normalization, FA, bi-clustering and feature components selection.

3.1 Density Normalization

Relevance is the core idea behind IR. This measurement is not a distance but a probability on content similarity. As Zhai *et al.* argued [11], relevance is closely associated with distribution density of documents in a collection. Normalizing feature distributions is therefore an effective method to enhance the discrimination between documents and a query. According to the hypothesis of uniform term distribution, we project a document collection to a new feature space, in which documents are sparsely distributed with equal distribution density. For the convenience, a discrete space (feature term space) is used. The extraction of a *feature term* is a projection from a multiple valued N-dimensional variable to an integer, *i.e.* clustering which assigns class labels to data samples. This projection can be symbolized as a function $\hat{f} : [0, K]^N \rightarrow \{0, 1, \ldots, M-1\} \sim \{0, 1\}^M$, where K denotes the range of a feature and M the number of classes. We regard these integers as *feature terms*. In one-dimensional case, $N = 1$.

For a collection D, the frequency of a feature term f_t is the times that a feature falls into a given value interval $t \in [0, M)$ (Equation 1).

$$f_t = |D_t|, D_t = \{d|\hat{f}(d) = t, d \in D\} \tag{1}$$

where d is a document in D. The probability of a *feature term* t is,

$$p(t) = \frac{f_t}{\sum_{i=0}^{M-1} f_i} \tag{2}$$

There are many approaches available to complete the projection from a feature to feature terms and is compared in [9]. We propose the usage of maximized information entropy because of the robustness.

$$Entropy_s(M) = -\frac{1}{\sqrt{M-1}} \sum_{i=0}^{M-1} p(t_i) \log(p(t_i)) \tag{3}$$

After the computation of feature terms, we employ factor analysis to select discriminant components.

3.2 Factor Analysis

We randomly sample video frames from the media collection. A matrix F is therefore generated, in which each row contains feature terms from a visual frame. Factor analysis is applied on the covariance matrix of F which generates the loading matrix Λ.

3.3 Bi-clustering

We try three methods to cluster the loading matrix. The component clustering only considers the overall distance between two components. This distance measurement makes the similarity in factor patterns questionable, as we think about factor combinations as well. In factor clusters, we have a group of different factors which behave almost the same for all components. However, it will miss some factor combinations that behave similarly only for some components, due to the constraint that all objects in a cluster should contain all components. To overcome these problems, we turn to bi-clustering over the loading matrix. Bi-clustering is a two-way data analysis and aims to find subgroups of rows and columns, which are as similar as possible to each other and as different as possible to the rest. The BiMax algorithm [7] is used for bi-clustering the loading matrix Λ.

3.4 Communality Based Feature Selection

We select the component with minimum communality from each bi-cluster for efficient representation. This is because components with minimum communality have minimum variance in common with other feature components and are therefore more discriminative than other components.

4 Experiment and Results

The TRECVid 2006 collection is used for evaluation, including 160 hours news videos and 24 content-based queries (Topic 173-196). For each query topic, from seven to eleven images are used as query examples and a ground truth is provided as a ranked list of 65 to 775 relevant shots. We use the state-of-art of PCA-based K-mean clustering [5] as the baseline. Results from the baseline are denoted by Run 1 in Table.1.

Table 1. rel_ret@1000(r_r@1000), MAP and P@20 for different experiments

		80% Feature Selected						60% Feature Selected					
		r_r@1000		MAP		P@20		r_r@1000		MAP		P@20	
RUN	Description	EH	HT	EH	HT	EH	HT	EH	HT	EH	HT	EH	HT
1	PCA + KMeans Centers(Baseline)	331	160	.0043	.0007	.0417	.0093	340	133	.0057	.0003	.0561	.0063
2	All feature components	384	199	.0066	.0009	**.0625**	**.0146**	same	bec.	no	feat.	sel.	
3	FA + Bi-Clust + Max. Comm.	335	162	.0059	.0004	.0458	.0125	331	123	.0067	.0005	.0583	.0063
4	FA + Bi-Clust Centers	360	170	.0074	.0005	.0500	.0083	347	135	.0057	.0003	.0396	.0063
5	FA + Bi-Clust + Min. Comm.	**389**	208	.0083	**.0012**	**.0625**	.0104	351	199	.0074	.0010	.0583	.0125
6	Density Norm. + FA + Bi-Clust + Min. Comm. (Proposed)	382	**256**	**.0109**	.0010	**.0625**	.0042	345	**256**	**.0100**	**.0013**	**.0646**	.0104

Density normalization is carried out on the entire TRECVid 2006 collection. Keyframes are sampled every ten visual frames and we compute feature distribution across the sample collection. Since, factor analysis is of high computational complexity, we randomly selected 100 frames from the collection for factor analysis and bi-clustering.

Two MPEG-7 visual features, edge histogram (EH) (80 components) and homogeneous texture (HT) (62 components) are extracted, as both of features are of high dimensionality. The number of components are decided by the number of bi-clusters, as we select one component from each bi-cluster. In addition, the number of bi-clusters can be changed by users. For the convenience, we fixed the number of selected components to a given ratio of the original size, *i.e.* 80% and 60% respectively. The Euclidean distance between feature terms is used to calculate the dissimilarity between query examples and keyframes. The top 1000 shots that are of minimum distance from any query example will be returned as query results.

In Table 1, six runs are stated: Run-1 represents the baseline, Run-2 uses all feature components without any component selection; Run-3, Run-4 Run-5 and Run-6 are experiments with factor analysis and bi-clustering, but with different configurations. Run-3 tests the maximum communality components in bi-clusters; Run-4 uses components nearest to bi-cluster center; Run-5 selects the minimum communality components from each bi-cluster. Run-3/4/5 show the effectiveness of factor analysis and bi-clustering in feature subspace learning but work on the original low-level features, that is without density normalization. Run-6 is our proposed approach which combines density normalization, factor analysis, bi-clustering and minimum communality based component selection. Run-6 proves the effectiveness of density normalization. Run-1 and 4 highlight the performance difference between PCA and FA based methods. In all configurations, both of low-level feature and both of the given ratio of components, factor analysis performs better than PCA. Run-3 and 5 verify the assumption that component with minimum communality is the most discriminating component in a bi-cluster. Experimental results strongly supports this assumption, as Run-5 significantly outperforms Run-3.

5 Conclusion

In this paper, we propose a statistical strategy to facilitate feature subset selection. The highlight of this work is the exploitation of the hypothesis from statistical information retrieval, which adapt a traditional feature selection scheme to the application of video retrieval. Experimental results show that our approach outperforms PCA-based K-mean clustering.

Acknowledgments

This research was supported by the EC under contract SALERO (FP6-027122) and SEMEDIA (FP6-045032).

References

1. Amati, G., Rijsbergen, C.J.V.: Probabilistic models of information retrieval based on measuring the divergence from randomness. ACM Transactions on Information Systems (TOIS) 20(4), 357–389 (2002)
2. Collins, R., Liu, Y., Leordeanu, M.: Online selection of discriminative tracking features. IEEE Trans. PAMI 27(10), 1631–1643 (2005)
3. Harter, S.: A probabilistic approach to automatic keyword indexing, part i on the distribution of speciality words in a technical literature. Journal of the ASIS 26, 197–216 (1975)
4. Jiang, W., Er, G., Dai, Q., Gu, J.: Similarity-based online feature selection in content-based image retrieval. IEEE Transactions on Image Processing 15(3), 702–712 (2006)
5. Jolliffe, I.: Principal Component Analysis. Springer, New York (1986)
6. Lin, T.-S., Meador, J.: Statistical feature extraction and selection for ic test pattern analysis. IEEE International Symposium on Circuits and Systems 1, 391–394 (1992)
7. Madeira, S., Oliveira, A.: Biclustering algorithms for biological data analysis: a survey. IEEE/ACM Transactions on Computational Biology and Bioinformatics 1(1), 24–45 (2004)
8. Margulis, E.: N-poisson document modelling. In: SIGIR 1992, pp. 177–189. ACM Press, New York (1992)
9. Ren, R., Jose, J.M.: Query generation from multiple media examples. In: 7th International Workshop on Content-Based Multimedia Indexing, pp. 138–143 (2009)
10. Wang, J.: Mean-variance analysis: A new document ranking theory in information retrieval. In: Boughanem, M., et al. (eds.) ECIR 2009. LNCS, pp. 4–16. Springer, Heidelberg (2009)
11. Zhai, C., Lafferty, J.: A risk minimization framework for information retrieval. Inf. Process. Manage. 42(1), 31–55 (2006)