Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 6.017

IJCSMC, Vol. 6, Issue. 3, March 2017, pg.178 – 183

Enabling Fully Automated Learning for Application Specific Web Videos

Nishikesh Dalkari, Amey Neve, Vedang Naik, Mansi Mudkanna
K.K.Wagh Institute of Engineering Education & Research, Savitribai Phule Pune University, Nashik
nishikeshdalkari@gmail.com, ameyneve@gmail.com, vedangnaik3339@gmail.com, mansimudkanna@gmail.com

Abstract— ***Technology is an essential part of our lives today. The growing popularity of Web videos for day-to-day requirements necessitates to classify Web videos into application-specific categories, since different applications are interested in different aspects of the user preferences. Personalization applications need to understand user preferences in order to provide customized services. As user engagement with web videos has grown significantly, understanding user preferences based on videos viewed looks promising. This requires ability to classify web videos into a set of categories appropriate for the user application. Implementation a fully automated framework to obtain training videos to enable classification of web videos to any arbitrary set of categories, as desired by the user application.***
Keywords— ***Feature Selection, Clustering, Environment-specific retrieval, Supervised learning, Discrete Optimisation, Simulating Annealing, Training, Metadata, Classification, Video retrieval, Linear Proximity, Linear Diversity, Candidate keywords***

INTRODUCTION

We propose a fully automated framework to obtain training videos with properties that can lead to high performance of trained classification models. To achieve the above, the proposed framework neither relies on labels associated with online videos, nor requires any manual labelling of videos. Instead, we develop approaches to select keywords based on their suitability to retrieve high quality training videos for a specified set of categories.

In order to select the keywords, we first present a tunable approach (LCPD) based on the Linear Combination of Proximity and Diversity. While such an approach can give good classification performance, it requires the tuning of a parameter in order to obtain optimal performance, which may require manual effort. We thus also propose an approach using Annealing based Alternating Optimization (AAO) where we balance between the two objectives such that the final solution is a trade-off between the two. Complexity, convergence and correctness of the proposed algorithms are presented, along with experiments over several sets of categories.

Over the past few years, there has been a steady rise in the number and popularity of personalization applications available on the Internet. These include applications based on personalized advertisements, content recommendation systems, social network connection suggestions, and several others that attempt to understand the preferences of users.
Personalization applications have been traditionally based on learning user preferences through queried keywords and viewed articles. The last few years have also witnessed a tremendous increase in viewing and

sharing of web videos (such as on YouTube), with significant increases in unique viewers, total streams viewed, number of streams per viewer, and the time per viewer. Given their unique characteristics, web videos offer a tremendous potential for understanding user preferences.

User preferences can be inferred based on the types/ categories of web videos seen. Such videos are generally organized at video sharing websites on the basis of labels that the video uploaders choose from among a set of common categories that are used by such websites. Examples of such common categories include Comedy, Music, People, Entertainment, Pets, Science, etc. On the other hand, the categories of interest to personalization applications may be arbitrary, and quite different from the above common categories.

Consider a department store (such as Sears or Walmart) that might want to offer promotional coupons to buyers. Knowing whether a person (a buyer) has interest in product specific categories like fitness equipment, clothing items, or baby products would be of high interest to the department store, as compared to knowing whether he/she is interested in the common categories mentioned above. A movie recommendation system would like to learn if a viewer prefers action, horror, or comedy movies. Categorizing viewed videos and understanding user preferences in terms of the common categories used by video sharing websites might not be useful for different personalization applications. In addition to the above observation, it should be noted that different personalization applications are interested in understanding user preferences with respect to very different sets of categories, as shown by the above examples. It is clearly not sufficient to use a common set of categories for every personalization application, as the categories of interest for one application might be irrelevant and useless for another.

This calls for techniques to classify viewed web videos, and hence estimate user preferences, in terms of any arbitrary set of categories appropriate for a given personalization application. Various modes of information (such as audio, visual, textual and social network) can be employed to assist in the classification of web videos. Classifier employed for this task have the inherent requirement of training videos labelled to the set of categories as desired by the personalization application. Since the set of categories suitable for a personalization application might be very different from the common categories used by video sharing websites, training videos labelled according to the required set of categories are often unavailable. Our work addresses this requirement of obtaining training videos labelled as per the required set of categories, which are not necessarily the categories commonly associated with web videos. We propose a fully automated framework to obtain training videos with properties that can lead to high performance of trained classification models. To achieve the above, the proposed framework neither relies on labels associated with online videos, nor requires any manual labelling of videos. Instead, we develop approaches to select keywords based on their suitability to retrieve high quality training videos for a specified set of categories. Such a methodology requires the consideration of two opposing objective, namely proximity and diversity.

In order to select the keywords, we first present a tunable approach (LCPD) based on the Linear Combination of Proximity and Diversity. While such an approach can give good classification performance, it requires the tuning of a parameter in order to obtain optimal performance, which may require manual effort. We thus also propose an approach using Annealing based Alternating Optimization (AAO) where we balance between the two objectives such that the final solution is a trade-off between the two. Complexity, convergence and correctness of the proposed algorithms are presented, along with experiments over several sets of categories.

METHODOLOGY

[1]This paper was the first comprehensive study and large-scale test on web video (so-called user generated video or micro video) categorization. Observing that web videos are characterized by a much higher diversity of quality, subject, style, and genres compared with traditional video programs, author focus on studying the effectiveness of different modalities in dealing with such high variation. Specifically, author propose two novel modalities including a semantic modality and a surrounding text modality, as effective complements to most commonly used low-level features. The semantic modality includes three feature representations, i.e., concept histogram, visual word vector model and visual word Latent Semantic Analysis (LSA), while text modality includes the titles, descriptions and tags of web videos. Author conduct a set of comprehensive experiments for evaluating the effectiveness of the proposed feature representations over different classifiers such as Support Vector Machine (SVM), Gaussian Mixture Model (GMM) and Manifold Ranking (MR). Our experiments on a large-scale dataset with 11k web videos (nearly 450 hours in total) demonstrate that the proposed multimodal feature representation is effective for web video categorization and SVM outperforms GMM and MR on nearly all the modalities.

[2]In this paper the general problem of internet video categorization was examined. We make no assumptions about the videos we attempt to categorize: each video may be recorded from a hand-held video camera, a cellphone, a webcam, a television broadcast, or even an animated cartoon. An excellent source of such a wide variety of videos is the growing number of user-submitted video websites which have become popular over the last few years. As such, we explore the video categorization problem on two new databases of approximately 1000 and 1500 online user-submitted videos which we will be making available to the community.

[3]This paper presents a large scale video taxonomic classification scheme (with more than 1000 categories) tackling these issues. Taxonomic structure of categories is deployed in classifier training. To compensate for the lack of labeled video data, a novel method is proposed to adapt the web-text documents trained classifiers to video domain so that the availability of a large corpus of labeled text documents can be leveraged. Video content based features are integrated with text-based features to gain power in the case of degradation of one type of features. Evaluation on videos from hundreds of categories shows that the proposed algorithms generate significant performance improvement over text classifiers or classifiers trained using only video content based features.

[4]This paper presents a novel approach for web video categorization by leveraging Wikipedia categories(WikiCs) and open resources describing the same content as the video, i.e. Content-duplicated open resources(CDORs).Note that current approaches only collect CDORs within one or a few media forms and ignore CDORs of other forms. We explore all these resources by utilizing WikiCs and commercial search engines. Given a web video, its discriminative Wikipedia concepts are first identified and classified. Then a textual query is constructed and from which CDORs are collected. Based on these CDORs, we propose to categorize web videos in the space spanned by WikiCs rather than that spanned by raw tags. Experimental results demonstrate the effectiveness of both the proposed CDOR collection method and the WikiCs voting categorization algorithm. In addition, the categorization model built based on both WikiCs and CDORs achieves better performance compared with the models built based on only one of them as well as state-of-the-art approach.

[5]This paper focuses on classification on video genres of cartoons, movies, advertisements, news, and sports. It can be served for video organization, retrieval, etc. Based on the analysis on different video genres, we fuse video's time feature and color feature from shots together. Specifically, there are seven features including gradient and color features and each one could be an expert for some genre of video. We select these expert features and let them collaborate to improve the accuracy of classification. Then support vector machine (SVM) is used for classification. Experimental results on large amount of video demonstrate the effectiveness of the proposed method.

This paper describes the methodology of problem solving used is algorithmic approach. The algorithm used is Linear combination of proximity and Diversity (LPCD).

Linear combination of proximity and Diversity (LPCD):
Valid keywords are obtained by applying high proximity and high diversity. High proximity algorithm selects keywords which are closer to given Set of Categories High Diversity Algorithm eliminates keywords which are farthest from given set of Categories. Valid Categories are further given to Classifier which obtains Training Dataset.

Annealing based alternating optimization (AAO):
The diversity of distances measures between the keywords, however an optimization technique based on such a measure would essentially boil down to performing a combinatorial search over all subsets of the set of valid keywords.

ALGORITHM

In order to obtain SRKs from a set of valid keywords, we propose two efficient approaches. The first approach (LCPD) is based on defining a diversity score per keyword. By defining a linear combination of the proximity and diversity scores of keywords, we propose an iterative algorithm to select SRKs one by one. Such an approach leads to high accuracy in video classification based on varying a parameter that controls the relative importance given to the two objective functions, namely proximity score and diversity score of a selected keyword.

For category *i*, we define **Valid Candidate Keywords** (called valid keywords for brevity) as those Candidate Keywords that retrieve more videos having true label of category *i* than of any other category. Then for a candidate *K* of category *i*, *K* is a valid keyword if

$$|\mu_k - \mu_i| < |\mu_K - \mu_j| \;\square_j \neq i.$$

Here $\mu_K = (\Sigma v{:}v \epsilon RV(K)/|RV(K)|)$. The true mean *i* for category *I* can be approximated as centroid of *RV*(*Ci*), i.e., of the set of videos retrieved by name of category *i*.

In order to combine the Proximity and Diversity scores to obtain *SRT* score of a valid keyword *K*, we assume *SRT* to be a simplistic linear combination of the two scores. *SRT* (*K*; *T* 0(*i*)) denotes the Suitability for Retrieving Training video score of a valid keyword *K* for category *i*, given that existing training data for category *i* is *T* 0(*i*). This equation is called **Validity Filter**.
Only valid keywords should be considered for being selected as SRK to ensure more number of training videos are added in *T* (*i*) that have true label of category *i* than videos that are mislabelled as category *i*.

$$SRK\ (K, T'(i)) = \alpha * \{(N_1/|\mu_K - \mu_i|)\} + 1\ (1-\alpha)$$
$$*\{N_2.div(T(i)\ U\ RV(K))\}$$

Here, *N*1 and *N*2 are normalization factors used to ensure that Proximity and Diversity scores have the same order of magnitude. α ϵ [0; 1] is the **moderation factor**, which decides the weight given to the Proximity score relative to the Diversity score. We next discuss an iterative algorithm to obtain (a maximum of) *L* keywords as SRK from a given set of candidates for each category, using *SRT* as calculated above.

SRK Algorithm-
Inputs:
Names of categories: C_i
Number of SRKs: L
Candidate keywords per category: $K_{Cabdidates,i}$
Initialization:
$K_{SRk,i} \leftarrow$ [] (empty set)
$T'(i) \leftarrow RV(C_i)$
Applying Validity Filter:
$K_{valid,i} \leftarrow \{K \epsilon K_{Candidates,i}\}$: K satisfies (4)
Iterative Algorithm:
For n=1 to L
If $|K_{valid,i}|$ =0: STOP
Calculate $SRT(K, T'(i)) \square K \text{ e } K_{valid,i}$
$K_{top} \leftarrow \arg\max_K SRT(K,T'(i))$
$K_{SRK,i} \leftarrow K_{SRk,i}\ U\ K_{top}$
$T'(i) \leftarrow T'(i)\ U\ RV(K_{top})$
$K_{valid,i} \leftarrow K_{valid,i} \backslash K_{top}$
End For
Output: $K_{SRK,i}$

Assume that *M* Candidate Keywords are available for each category. Let the set of Candidate Keywords for category *i* be $K_{Candidates,i}$. For each category, a set of valid keywords $K_{Valid,i}$ is selected as a subset of $K_{Candidates,i}$ that satisfy. Starting with *T* 0(*i*) D *RV*(*Ci*), *SRT* (*K*; *T* 0(*i*)) is calculated for each valid keyword using, and the top keyword K_{top} is selected as an SRK. *T* 0(*i*) is then updated to *T* 0(*i*) [*RV*(K_{top}), and the process is repeated until *L* SRKs are selected or there are no valid keywords left. The proposed algorithm selects SRKs in an iterative manner, as compared to ranking valid keywords by their *SRT* score calculated once and selecting the top *L* keywords. While the latter calculates *SRT* scores independent of other SRKs selected, the proposed algorithm attempts to increase Intra-Category Diversity of the resulting training data, leading to better performance of trained classification model.

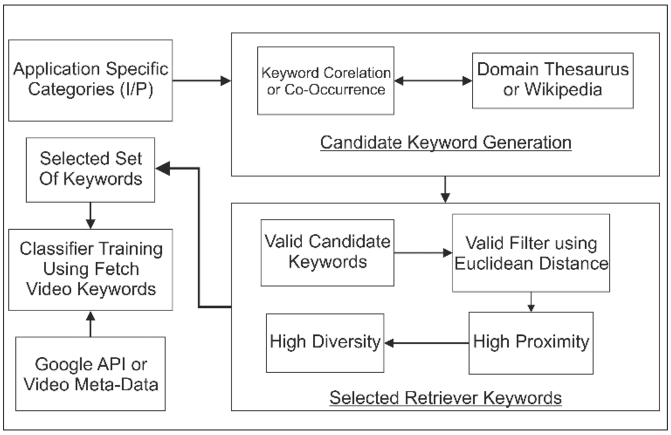

Fig. 1: Steps to obtain Selected Retriever Keyword (SRK)

Description:

Application Specific Category is given as input. Candidate Keyword is generated using Domain Thesaurus or Wikipedia (Database). From Candidate Keywords, Valid Candidate Keywords is selected. These Keywords are selected by Validity Filter using Euclidian distance between Feature Vectors of the Valid Candidate keywords. Further Precise selection of keywords is done by LCPD (Linear Combination of Proximity and Diversity), in this, Proximity finds the most precise keywords and Diversity eliminates distant Keywords. Classifier is trained using Selected Set Keywords and Videos are fetched precisely using classifier.

CONCLUSION

The system is a user personalization search using categories and keyword. This system will give precise and related videos based on the requirement of user. We have proposed a fully-automated framework to obtain high quality training videos for any arbitrary set of categories, without the need for any manual labelling that is needed by most related approaches. We analyze properties of training data that lead to high performance of the trained classifier. Based on the above properties, we propose approaches for selecting keywords to retrieve training videos, on the basis of their proximity to the categories of interest, and the resulting diversity in the training data. The first approach (LCPD) leads to high classification accuracy, although requires a parameter representing preference for defined objective functions.

The parameter can be obtained through manually provided preferences for the objective functions, or by using manually labelled validation videos for tuning. In order to avoid the manual effort, we also provide an annealing based alternating optimization framework (AAO) and propose its adaptive variant (Adapt AAO) to select keywords. Such an approach does not require the articulation of preferences in parameterized form, with the trade-off of some loss in classification accuracy as compared to the LCPD approach. Experimental results on several sets of categories show the effectiveness of the training videos obtained by the proposed approaches, hence making classification of videos watched by users to arbitrary set of categories feasible. Consequently, this work may enable new personalization applications by enabling identification of user preferences in a set of categories relevant to the application.

REFERENCES

[1] L. Yang, J. Liu, X. Yang, and X.-S. Hua, ``Multi-modality Web video categorization,'' in *Proc. ACM Int. Workshop Multimedia Inform. Retrieval.* , 2007, pp. 265_274.

[2] G. Schindler, L. Zitnick, and M. Brown, ``Internet video category recognition,'' in *Proc. IEEE Computer. Soc. Conf. Computer. Vis. Pattern Recognition. Workshops*, Jun. 2008, pp. 1_7.

[3] Y. Song, M. Zhao, J. Yagnik, and X. Wu, ``Taxonomic classification for Web-based videos,'' in *Proc. IEEE Conf. Computer. Vis. Pattern Recognition.*, Jun. 2010, pp. 871_878.

[4] Z. Chen, J. Cao, Y. Song, Y. Zhang, and J. Li, ``Web video categorization based on Wikipedia categories and content-duplicated open resources," in *Proc. ACM Int. Conf. Multimedia*, 2010, pp. 1107_1110.
[5] J.-F. Chen, H. Lu, R. Wei, C. Jin, and X. Xue, ``an effective method for video genre classification," in *Proc. ACM Int. Conf. Image Video Retrieval.*, 2010, pp. 97_104.
[6] J. R. Zhang, Y. Song, and T. Leung, ``Improving video classification via
YouTube video co-watch data," in *Proc. ACM Workshop Soc. Behavioral Network, Media Access*, 2011, pp. 21_26.
[7] C. Verma and S. Dey, ``fully automated learning for application-specific Web video classification," in *Proc. IEEE/WIC/ACM Int. Joint Conf. Web Intelligence, Intelligence. Agent Technol.*, Nov. 2013, pp. 307_314.
[8] C. E. Brodley and M. A. Friedl, ``Identifying and eliminating mislabelled training instances," in *Proc. Nat. Conf. Artificial. Intelligence*, 1996, pp. 799_805.
[9] N. Shroff, P. Turaga, and R. Chellappa, ``Video précis: Highlighting diverse aspects of videos," *IEEE Trans. Multimedia*, vol. 12, no. 8, pp. 853_868, Dec. 2010.
[10] *YouTube*. [Online]. Available: http://www.youtube.com, accessed Jan. 2017.
[11] *Vimeo*. [Online]. Available: http://www.vimeo.com, accessed Jan. 2015.
[12] *Reverse Dictionary*. [Online].Available: http://www.onelook.com/reversedictionary.shtml, accessed Jan. 2017.
[13] G. Golub, ``Numerical methods for solving linear least squares problems," *Numerical Math.*, vol. 7, no. 3, pp. 206_216, 1965.
[14] X. Yang, Q. Song, and Y. Wang, ``A weighted support vector machine for data classification," *Int. J. Pattern Recognition. Artificial. Intelligence*, vol. 21, no. 5, pp. 961_976, 2007.

