A Multi Faceted Recommendation Approach for Explorative Video Retrieval Tasks

David Vallet, Martin Halvey, David Hannah and Joemon M. Jose

Department of Computing Science
University of Glasgow
{dvallet, halvey, davidh, jj}@dcs.gla.ac.uk

ABSTRACT

In this paper we examine the use of multi faceted recommendations to aid users while carrying out exploratory video retrieval tasks. These recommendations are integrated into ViGOR (Video Grouping, Organisation and Retrieval), a system which employs grouping techniques to facilitate video retrieval tasks. Two types of recommendations based on past usage history are utilised, the first attempts to couple the multi-faceted nature of explorative video retrieval tasks with the current user interests in order to provide global recommendations, while the second exploits the organisational features of ViGOR in order to provide recommendations based on a specific aspect of the user's task.

Author Keywords: Video, collaborative, recommendation, exploratory, search.

ACM Classification Keywords: H.5.1 Multimedia Information Systems, H.5.3 Group and Organization Interfaces

General Terms: Design, Experimentation, Human Factors

INTRODUCTION

Current state of the art systems that are used to organise and retrieve video are insufficient for dealing with the vast and swiftly growing volumes of video that are currently being created. Specifically, there is a growing need to create tools and techniques to aid users in the difficult task of searching for video; this is particularly true online with the increasing growth of web-based video storage and search systems. Current video retrieval systems rely on textual descriptions or methods that use low-level descriptors to find relevant videos. Neither of these methods has proved to be sufficient to overcome the difficulties associated with video search. The difference between the low-level data representation of videos and the high level concepts that people associate with video, commonly known as the semantic gap, provides difficulties for using these low-level features. We propose that many of the problems associated with searching large collections of video can be eased through the use of recommendation techniques. Recommendation techniques offer a solution that allows systems to circumnavigate the problems associated with the semantic gap [3] and the unreliability of textual descriptions.

To that end, we have extended our existing recommendation techniques that exploit both the implicit and explicit actions involved in previous user searches [3]. This extension in itself is an innovation as it allows us to provide recommendations that support complex and explorative search tasks. However, the main goal of this work is to assist users in completing their difficult search tasks, by creating a predictive model that exploits both the implicit and explicit actions involved in previous user searches to provide multi-faceted recommendations. To achieve this we provide recommendations to users using two novel techniques: 1) a global recommendation that relates to the overall goal of the search task that they are carrying out and 2) a local recommendation that relates to the particular aspect of a search task that users are exploring at that time. We believe that our approach of modelling multiple aspects of user needs via implicit user interactions can result in improved user performance in terms of task completion and reduce the user effort involved in finding relevant videos.

SYSTEM DESCRIPTION

Figure 1 shows the ViGOR interface, ViGOR comprises of a search panel (A), results display area (B), workspace (C) and playback panel (D). The users enter a text based query in the search panel to begin their search. The result panel is where users can view the search results (b). Additional information about each video shot can be easily retrieved. Placing the mouse cursor over a video keyframe for longer than 3 seconds will result in any text associated with that video being displayed to the user (we will hence forth refer to this action as tooltip) (f). Users can play, pause, stop and navigate through the video as they can on a normal media player. Like the MediaGLOW [1] and EGO [4] systems, the most important element of ViGOR is the provision of a workspace (C). In MediaGLOW and EGO the workspace is only used to cluster images, however as has been discussed previously [2] the difficulties relating to video and image search are somewhat different and the approach of using groups in a workspace is an extremely useful solution for

IUI'10, February 7–10, 2010, Hong Kong, China.

video search. Groups can be created by clicking on the create group button. Users must then select a textual label for the group. Drag-and-drop techniques allow the user to drag videos into a group or reposition the group in the workspace. Any video can belong to multiple groups simultaneously. The workspace is designed as a potentially ever expanding space to accommodate a large number of groups. Each group can also be used as a starting point for further search queries. Users can select particular videos and can choose to view an expansion of the group that contains similar videos based on a number of different features (d). As the ViGOR system uses the YouTube API as a backend, the features available to expand the group are mainly standard YouTube features. The interface offers three expansion options (e): 1) related videos; 2) videos from the same user 3) and text expansion, which is the result of a new search using text extracted from the selected videos.

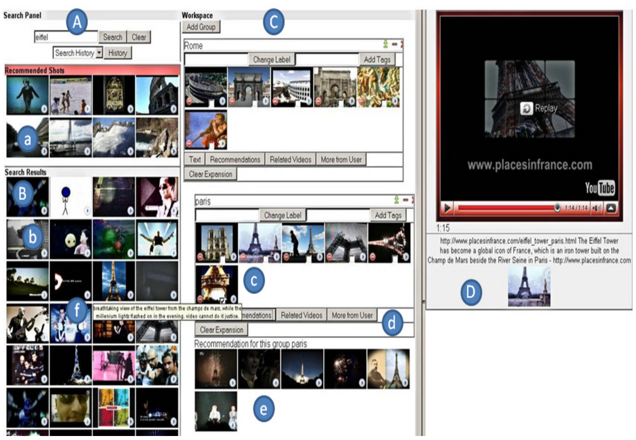

Figure 1. Interface of the video retrieval system

Extending ViGOR with recommendations

Some of the interface components of ViGOR allow users of the system to provide explicit and implicit feedback, which is then used to provide recommendations to future users. Explicit feedback is given by users adding a video to a group (c). Implicit feedback is given by users playing a video (D), highlighting a video using the tooltip (f) or submitting a search query (A). In this extended interface, users also have two options to receive recommendations. The users are presented with global recommendations of video shots that might match their search criteria based on their interactions (a). Users may also retrieve localised recommended videos via an extra expansion option available in the group (d, e). These recommendations are localised to each group and are based on the interactions of previous users with videos that the current user has selected.

A MULTI-FACETED RECOMMENDATION APPROACH

The goal of our recommendation approach is twofold: 1) to exploit the organisational functionalities provided by ViGOR as a new source of implicit and explicit information; and 2) to take into consideration the ambiguous and multi-faceted nature of explorative video tasks. We follow a graph-based approach for the representation of past user interactions, based on previous work [3]. In this approach, a user session s is represented as a set of queries $\mathrm{Q_s}$, which were input by the user, the set of multimedia documents $\mathrm{D_s}$ the user accessed during the session, and a set of groups or aspects $\mathrm{G_s}$ the user created during the search session.

Queries, documents and groups are the nodes $\mathrm{N_s} = \mathrm{Q_s} \cup \mathrm{D_s} \cup \mathrm{G_s}$ of our graph representation $\mathrm{G_s} = (\mathrm{N_s}, \mathrm{W_s})$. The arcs of this graph representation, $\mathrm{W_s}$, are of the form $(\mathrm{n_i}, \mathrm{n_j}, \mathrm{u}, \mathrm{w_s})$ and indicate that at least one action led the user u from the node $\mathrm{n_i}$ to $\mathrm{n_j}$. Note that the only action that can lead to a group node $\mathrm{g} \in \mathrm{G_s}$ is the action of assigning a document node to the group. The weight value $\mathrm{w_s}$ represents the probability that node $\mathrm{n_j}$ was relevant to the user for the given session. This value is either given explicitly by the user, or estimated by means of the implicit evidence obtained from the interactions of the user with that node, following a previously developed implicit model [3].

Finally, all the session graphs are aggregated into a single graph $\mathrm{G} = (\mathrm{N}, \mathrm{W})$, where $\mathrm{N} = \bigcup_s \mathrm{N_s}$ and $\mathrm{W} = \bigcup_s \mathrm{W_s}$, which constitutes a global pool of usage information that collects all the implicit and explicit relevance evidence of users from past sessions. The nodes of the implicit pool are all the nodes involved in any past interaction $N = \bigcup_s N_s$, whereas the weighted links W are of the form (n_i, n_j, w), where $n_i, n_j \in N$ and w combines the probabilities of all the session-based values. As weight values are considered probabilities of relevance of the node n_j to the user, we opt for a simple aggregation of these probabilities, this is $w = \frac{\sum_S w_S}{|w_S|}$. Each link will thus represent the overall implicit (or explicit, if available) relevance that all users whom actions or soft links led from node n_i to n_j, gave to node n_j.

Global Recommendation

The global recommendation approach is based on the status of the current user session. As the user interacts with the system, a session graph $G_s = (N_s, W_s)$ is constructed, where in this case s is the current user's ongoing session. This graph is the input for the global recommendation algorithm presented next. This recommendation approach has the main goal of exploiting the implicit pool in order to retrieve similar nodes that were somewhat relevant to other users and. We define the global recommendation as:

$$gr(n, N_s) = \sum_{\substack{n_i \in N_s \\ p = n_i \rightsquigarrow n_j \to n \\ length(p) < D_{MAX}}} lr'(n_i) \cdot \xi^{length(p)-1} \cdot w\left(n_j, n\right)$$

Where $n_i \rightsquigarrow n_j$ denotes the existence of a path from n_i to n_j in the graph, taking link directionality into consideration. $n_j \to n$ means that n is adjacent to n_j. $w\left(n_j, n\right)$ is the probability weight, given by the implicit pool. $lr'(n_i) \in [0,1]$ is a weighing function that follows our previous implicit model [3] based on the relevance of this node to the outgoing user's session, obtained from the user's implicit and explicit feedback. *Length*(p) is counted as the number of links in path p, which must be less than a maximum length D_{MAX}, set to 5 in for our evaluation. Finally, ξ is a length reduction factor, set to 0.8, which allows us to give more importance to those documents that directly follow the interaction sequence.

Local Group Recommendation

The local group recommendation focuses on a single group, and tries to recommend more documents that could aid the user in expanding the aspect of the task represented by this group. This recommendation approach exploits the representation in the implicit pool of the different aspects created by previous users. In this case, the local group recommendation tries to find similar aspects that previous users could have created and then rank their related documents. The input of the local group recommendation is the set of documents $\mathrm{D_g}$ that the current user has selected within an aspect group $\mathrm{g} \in \mathrm{G}$. On the first step of this approach, related aspect groups from the implicit pool are searched for, this is achieved by ranking the related groups panels $n_g \in \mathrm{G}$ with the global recommendation approach, using the set of selected documents as input, and ranking only group nodes. Thus, the related groups can be ranked as $gr(n_g, D_g)$, were the local relevance of the selected documents is set to 1, i.e. $lr'(d_g) = 1, d_g \in D_g$. We limited the expansion distance D_{MAX} to 3, in order to constraint the search to more related groups.

On the second step of this approach, the implicit pool is exploited in order to rank the top nodes related to the set of ranked group nodes n_g. The ranking approach in this second step is to rank higher those documents that belong to more related aspects created by previous users. This ranking can be also implemented by tuning the global recommendation approach in the following way. The input of the ranking approach will be the ranked set of related groups, $n_g \in \mathrm{G}$. The local relevance of the input is the ranking given by the previous step: $lr'(n_g) = gr(n_g, D_g)$. Thus, the final ranking is obtained from $gr(n, n_g)$.

EXPERIMENTAL METHODOLOGY

In order to evaluate our multi-faceted approach for video recommendation a user evaluation was conducted, the goal of which was to answer a number of research questions.

R1 Does the user performance improve or diminish with ViGOR that provides recommendations in comparison with the same system without recommendations? Are these recommendation approaches effective enough?

R2 Does the provision of recommendations impact on the effort involved for users in searching and exploring a video collection?

R3 Do the users use different types of recommendations, i.e. global recommendations vs. local recommendations, at different stages of their task or is the use of these tools independent of the stage of the task?

Experimental Design

For the purposes of this evaluation, the YouTube API was used to provide access to YouTube's video collection. Four simulated work task situations were created in order to provide broad, ambiguous, open ended tasks for the users. The system provides recommendations based on logs from a previous evaluation of ViGOR that also used YouTube [2] and four simulated tasks. The tasks defined in this evaluation had some similarities. However, the new tasks for this evaluation are broader than the tasks for the previous evaluation and might contain only some aspects related to the tasks from the previous evaluation. The four evaluated tasks are:

Task 1 Politics: A task of finding videos containing leading world figures.

Task 2 Travel: A task of finding video clips about locations in Europe that you would like to visit.

Task 3 Culture: A task of finding videos that illustrate Scottish culture.

Task 4 News: A task of finding videos illustrating news stories from 2008.

A between subjects design was adopted for this evaluation. Two systems were evaluated; ViGOR with and without recommendations. Tasks and interfaces were assigned using a Latin square design. The participants were given ten minutes training on their search system carrying out training tasks. Users had a maximum of 20 minutes to complete each task. For each participant their interaction with the system was logged and the videos they marked as relevant were stored. 24 participants took part in our evaluation. The participants consisted of 18 males and 6 females with an average age of 28.78 years (median: 28) and an advanced proficiency with English. The participants indicated that they regularly interacted with and searched for multimedia. They were paid a sum of £12 for their participation, which took approximately 2 hours.

RESULTS

Task Performance

A direct comparison between the two interfaces found that on average users of the recommendation system marked 27.15 videos as being relevant (i.e. add them to groups) in comparison with 19.6 videos for users of the system without recommendations. This is an increase of 38.47% in the number of retrieved relevant videos. In addition to this, users of the recommendation system created more groups or aspects of the task on average, 5.4, as opposed to 4.67 for the system without recommendations, this is an increase of 15.63%. It was found that system was a significant variable for the number of videos retrieved (2 way ANOVA, $F=6.94$, $p=0.01$). Overall, these results show that users are retrieving more videos and examining more aspects of their task using ViGOR with recommendations.

A further analysis was performed on the interaction logs. Users of ViGOR with recommendations have more user interactions with the system overall in comparison with users of the ViGOR baseline system. However, much of this difference is due to the increased use of the tooltip functionality of the recommendation system; this is a lightweight functionality which is of low cost for the user to carry out. In terms of more heavyweight user actions such as querying the system or viewing a video, there are small differences between the two systems, with no statistically significant differences. One major noticeable difference in the user interactions is the way that the users use the expansion functionalities. In the recommendation system the three YouTube related expansions are used less frequently than in the baseline ViGOR system. This is to be expected as this system offers one more expansion option. However, the new recommendation expansion is used almost as frequently as the YouTube related expansion. Users appear to find the recommendations useful and exploit this resource. A more direct comparison of the recommendation techniques shows that the global recommendation techniques are selected as relevant more often than the local recommendations.

Thus far we have seen that the user performance improves with the use of the recommendations, this addresses the first of our research questions, as we were able to exploit past noisy implicit information to benefit the users in their explorative and multi-faceted tasks. It has also been shown that the user interactions increase while using the recommendation system, however most of this increase is due to an increase in the use of the lightweight tooltip function, this may just be as a result of the extra results and options that are presented to the users of the recommendation system. This addresses our second research question pertaining to user effort.

User Interactions

In an attempt to gain a further insight into the differences between the user interactions with the two types of recommendation approaches, we plotted a cumulative distribution for selecting each type of recommendation against time (see Figure 2). We do not show any of the other three expansions actions in this figure, as they follow a similar distribution to the local recommendation. A pair wise t-test revealed that the differences between the recommendation distributions were statistically significant. Users select examples from the global recommendations early in the task, it is not until later in the task that the users appear to select examples from the local expansions and add them to groups. This illustrates a difference in user behaviour, it appears that at the beginning users are more interested in the overall global task, but as the task progresses users become more interested in the details of each aspect, thus providing an answer for our third research question.

DISCUSSION AND CONCLUSIONS

In this paper we have presented our approach for providing multi faceted recommendations by extending our implicit feedback recommendation approach [3]. In order to evaluate this approach for providing recommendations and indeed the concept of multi-faceted recommendations for videos search, the recommendation algorithms were integrated into our ViGOR system. The unique organisational features available in ViGOR allow richer and multi-faceted recommendations.

The results of our evaluation have highlighted the promise of multi-faceted recommendations for video search tasks. Our recommendation approach based on implicit feedback coupled with an innovative video search interface has improved user performance and highlighted the promise of multi-faceted recommendations based on collaborative implicit feedback for alleviating many problems associated with online video search, and indeed could be applied to numerous other video search paradigms.

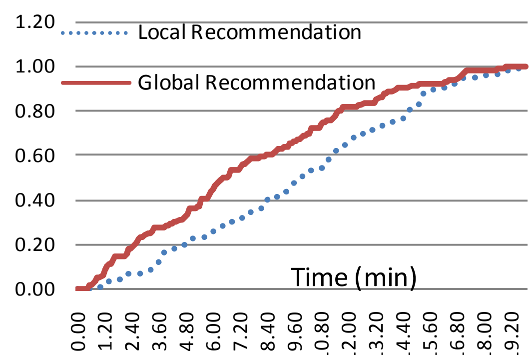

Figure 2. Cumulative distribution of selection of recommendations over time

Acknowledgements. This research was supported by the European Commission under contract FP6-027122-SALERO and by the Spanish Ministry of Science and Education (TIN2008-06566-C04-02).

REFERENCES

1. Girgensohn, A., Shipman, F., Wilcox, L., Turner, T., and Cooper, M. MediaGLOW: organizing photos in a graph-based workspace. In *Proc.* IUI 2009, 419 - 424.
2. Halvey, M., Vallet, D., Hannah, D., and Jose, J. M. ViGOR: a grouping oriented interface for search and retrieval in video libraries. In *Proc. JCDL 2009*, 87-96.
3. Hopfgartner, F., Vallet, D., Halvey, M., and Jose, J. Search trails using user feedback to improve video search. In *Proc. ACM MM 2008*, 339-348.
4. Urban, J. and Jose, J.M. A Personalised Multimedia Management and Retrieval Tool. In the International Journal of Intelligent Systems, 21(7), 725-745, 2006.