Abductive retrieval for multimedia information seeking

Ian Ruthven[1], Mounia Lalmas[2] and Thomas Rölleke[2]

[1]Department of Computer and Information Sciences, University of Strathclyde, Glasgow G1 1XH. Ian.Ruthven@cis.strath.ac.uk.
[2]Department of Computer Science, Queen Mary, University of London, London E1 4NS. mounia,thor@dcs.qmul.ac.uk

Abstract

In this paper we discuss an approach to the retrieval of data annotated using the MPEG-7 multimedia description schema. In particular we describe a framework for the retrieval of annotated video samples that is based on principles from the area of abductive reasoning.

1 Introduction

One of the most popular features of modern computing technology is that it allows almost any computer user to utilize a wide range of text, image, sounds, and video data. Unfortunately, *finding* these multimedia resources is often difficult.

Designing tools that provide easy access to multimedia data involves two main themes: defining methods to *describe* the data and providing access methods to *search* the data. The first theme – data description – is being tackled by the new standard **MPEG-7**, formally called "*Multimedia Content Description Interface*", which consists of a standard set of tools for describing the content of multimedia data [7]. The second theme, access methods, is the particular focus of our research.

MPEG-7 annotated multimedia data, such as video or speech, requires advanced search tools as MPEG-7 supports different levels of description; for example, the *structural* level (e.g. "this video consists of a sequence of segments and each segment is composed of several shots"), the *feature* level (e.g. "this object has the form of a flower"), and the *semantic* level (e.g. "Senegal beats France in the 2002 World Cup").

To fully exploit the variety of descriptions in the MPEG-7 representation what is required is a system that can utilise and reason about the different levels of description. Our proposed methodology is based on the notion of *abductive* explanation [4]. Given a small subset of data that the user has identified as being of interest, this form of knowledge manipulation generates *explanations* of why the data may be relevant to a searcher. An explanation is a description of the data containing elements from any of the MPEG-7 description layers associated with the data.

These explanations can be used in several ways. Firstly, by using the explanation as a new query, explanations can be used to facilitate the retrieval of additional relevant material. Secondly, explanations can be used to help the users understand how the retrieval system operates. Finally, explanations can be used to personalised the interaction with the user as the explanations are based on the individual user's interaction with the system. In this paper we provide an overview of our approach. We describe this approach specifically in relation to video data.

2 Searching MPEG-7 annotated data

MPEG-7 is a new standard for describing the content of multimedia data such as text, image, video and speech. MPEG-7 is restricted to a few, but powerful, concepts: a set of *descriptors* for representing features of multimedia material, a set of *description schemes* which define the structure and the semantics of the data, and a *description definition language* to specify descriptors and description schemes.

MPEG-7 allows a range of descriptions to be applied to a single object, from low-level content elements, through information related to the use of the object, to high-level conceptual elements [7]. MPEG-7, then, provides rich descriptions on which complex retrieval algorithms can be constructed[1]. However, the complexity and variety of the MPEG-7 descriptions can make it difficult for *users* to specify what kind of objects they want to retrieve.

Three main approaches have been suggested to facilitate searching MPEG-7 annotated video data:

a) *Form-based* approaches, e.g. [1], in which users can enter query concepts for any of the MPEG-7 descriptions. For video retrieval users can specify aspects such as camera angle, colours used in the video, captions, or video duration. Once users have viewed some retrieved videos they can refine their query by altering the concepts used in the query. These very detailed approaches to querying are useful in situations where the user has a good understanding of what kinds of video they require before they start searching. However, if a user does not know what videos are available for retrieval, or if they are unsure of what kind of videos they want retrieved, then form-based approaches can be difficult to use successfully [1].
b) *Browsing*, surveyed in [6], in which users are presented with a relatively large number of video segments. Each segment may also be linked to similar segments. These approaches have the advantage that the user can quickly scan videos. However, these systems provide little or no support for the user in refining or focussing their search. Browsing systems also give the user little help in understanding *how* the system selects which videos to display, resulting in a loss of control over the search.
c) *Query-by-example* in which a user identifies examples of the type of videos required – the *relevant* videos - and the system uses the relevant videos to generate a new query with which to search the database. Query-by-example is potentially very powerful as, after the initial search, users do not need to *describe* what information they require, they only need to be able to *recognise* relevant items. However, so far, query-by-example systems have mainly concentrated on low-level features, such as texture, [13], or have only been used for specific functions, e.g. [14] used query-by-example to identify individual actors within video clips.

For MPEG-7 data a more sophisticated approach than simple query-by-example is necessary to exploit the multiple description levels and to combine MPEG-7 elements (descriptors and description schemes) from each level. What we propose in this research is an integrated approach to the retrieval and presentation of search results; an approach that reduces the cognitive effort a user must expend in describing what information they require. We base this approach on the notion of *abductive inference* [4].

[1] A number of these are surveyed in [3].

3 Abductive inference

Abductive inference is specifically designed to provide *explanations* of complex data. In our framework the data to be explained are the relevant video segments selected by the user. From this set of video segments an abductive inference system will generate a set of possible explanations of why these samples may have been assessed relevant. Each explanation is formed from the MPEG-7 annotations of the relevant video samples. The explanations can be formed from any level of the MPEG-7 annotation, and includes information on *what* descriptions elements are important for retrieval and *why* the elements are important.

Each explanation is a possible new query that can be submitted to the system to retrieve a new set of video segments. Information on what description elements are to be used specify the *content* of the new query, and information on why the elements are important specify how the elements are to be used to retrieve new video segments. In previous research we used the notion of abductive explanation to generate a framework for the retrieval of free-text documents [8-11]. Currently, we are using similar principles to develop an integrated model for structured MPEG-7 annotated data. In the following sections we describe how we represent information for retrieval, section 3.1, and how we use these representations to compose abductive explanations, section 3.2, and how we present explanations to the searcher, section 3.3.

3.1 Elements of explanations

MPEG7 allows rich and varied descriptions of a single object. However, these descriptions describe *what* information an object represents; they do not dictate *how* this information should be used to retrieve objects. Hence it is necessary to define representations of the MPEG-7 annotations in a form that will allow effective retrieval and query modification. Our approach is driven by a conceptual distinction between content, factual and contextual information. Some MPEG-7 descriptors are considered to represent content information (e.g. manually selected keywords), others to represent factual information (e.g. type of video). Relationships between descriptors (e.g., are descriptors are in the same description scheme) are treated as contextual information [2].

One of the main achievements of our previous work was to show that *multiple* descriptions of individual elements could lead to better retrieval results [9]. Multiple descriptions not only allow for more expressive retrieval (i.e. we can target which descriptions are important) but also more accurate retrieval (i.e. we only use those descriptions that are useful in classifying objects for individual searches). Therefore, for each description element, we have several methods of describing how an element is used in describing an object.

3.2 Composing explanations

Our approach to video retrieval is an advanced form of query-by-example; once the user has provided a sample of relevant data, abductive inference algorithms will be used to modify the user's query. This approach falls into three tasks; deciding what type of explanation is required, section 3.2.1, selecting description elements to create an explanation, section 3.2.2, and deciding how the explanation should be used to retrieve a new set of video segments, section 3.2.3.

3.2.1 Types of explanation

In our previous research, [11], we demonstrated that different types of query creation are more appropriate for individual searches. For example better retrieval effectiveness may be achieved by

increasing the range of concepts in the query, by only concentrating on high quality concepts or by concentrating on closely related concepts. These correspond to different *types* of explanation. The first step of the abductive inference algorithm is to decide what type of explanation is required for the current search. In our previous research we investigated using the subject's interaction with system to detect what kind of query modification was required [11]. A similar approach is used here to differentiate, for example, between searches where a user wants *any* video that mentions football or only videos in which *every* segment contains information on football. With our approach, therefore, we are selecting the retrieval algorithm itself based on the user's selection of relevant items.

3.2.2 Components of explanations

Once we have decided what type of explanation is required we must decide on what description elements are to be used to compose the explanation. That is, we select, based on the elements derived in 3.1, what elements are good at *discriminating* the relevant video segments from non-relevant segments. These elements are ones that are useful for retrieving objects that are similar to the ones liked by the user and can come from any of the MPEG7 description levels. How many elements are chosen, and the specific criteria for selecting elements, are dictated by the type of explanation required, as identified in 3.2.1.

3.2.3 How components are used in explanations

The above steps give us an explanation; a set of description elements from different description levels that are useful for a particular type of explanation. The final step is to decide how these elements should be used to retrieve a new set of video segments. Essentially, we ask in what way the description elements explain the relevant video segments; what aspect of the element leads to segments being assessed relevant. This decision is made based on the multiple descriptions of elements, and their relationships, given in section 3.1. That is, we not only elicit what *elements* are important but we also consider what descriptions of that *element* are important.

The overall model is based on a logical formalism, with associated probabilistic extensions. Logical reasoning is one of the core methods of modelling abductive processes [7].

3.3 Presenting explanations

One of the main aims of our research is to make the retrieval of MPEG-7 data more intuitive to the user. Abductive explanations can also be used at the interface level to help users search. Previously we have shown that explanation-based approaches are effective for textual retrieval in helping users understand why individual system decisions have been made, increasing user satisfaction with the system. Explanations also encourage searchers to interact with retrieval modification techniques. This latter finding is important since users are often reluctant to *trust* systems that automatically modify queries [8].

4 Summary

In this paper we outline current research on the interactive retrieval of multimedia data annotated using MPEG_7. Our approach is based on retrieval through abductive inference. This approach has the following advantages:

a) *Retrieval complexity is hidden*. It is easier for the system to reason about complex data than the user. By using abductive approaches the system can make complex retrieval decisions based on elements that the user may find difficult to manipulate, e.g. colour distribution, or to remember exactly, e.g. creation information. Explanations can also be complex, with components from all MPEG-7 levels; queries that the user would find it difficult to construct manually.

b) *Explanations have quality*. An abductive inference algorithm can create many possible explanations for the same set of data. Some of these explanations will be better descriptions of the data than others. However the quality of an explanation is strongly dependent on the *purpose* for which an explanation is required. For example some explanations would be more suitable for retrieving a small set of very similar video clips (a very precise search) than for retrieving *any* video clips that satisfy some criteria (a general background search).

c) *Explanations can aid understanding*. One of the attractive attributes of explanations is that they can be used to help users understand how search engines work; what features they use to retrieve video segments, how the search engines use these features and what changes the inference algorithm makes to a user's search. Therefore we can present the user with *reasons* for why their query was modified by the system and why new video segments were retrieved.

References

1. M.D. Dunlop and K. McDonald, *Supporting different search strategies in a video query interface*. Proceedings of RIAO 2000. Paris. 2000.
2. A. Graves and M. Lalmas. *Video retrieval using an MPEG-7 based inference network*. Proceedings of 25th ACM SIGIR. 2002.
3. J. Hunter. *MPEG-7 behind the scenes*. D-Lib Magazine. 5. 9. 1999.
4. J.R. Josephson and S.G. Josephson (ed). *Abductive Inference: Computation, Philosophy, Technology*. Cambridge University Press. 1994.
5. H. Lee and A. Smeaton. *Designing the user-interface for the Físchlár digital video library*. Journal of Digital Information. 2. 4. 2002.
6. P. Lipton. *Inference to the best explanation.* Routledge. 1991.
7. J. M. Martinez. *Coding of moving pictures and audio. MPEG-7 overview (v.8).* International organisation for standardisation. ISO/IEC. 2000.
8. I. Ruthven. *On the use of explanations as a mediating device for relevance feedback.* Proceedings of 6th ECDL Conference. Rome. 2002.
9. I. Ruthven, M. Lalmas and C.J. van Rijsbergen. *Combining and selecting characteristics of information use*. Journal of the American Society for Information Science and Technology. 53. 5. 2002.
10. I. Ruthven, M. Lalmas and C.J. van Rijsbergen. *Incorporating user search behaviour into relevance feedback*. Journal of the American Society for Information Science and Technology. *to appear*.
11. I. Ruthven, M. Lalmas and C.J. van Rijsbergen. *Empirical investigations on query modification using abductive explanations.* Proceedings of 24th ACM SIGIR. 2001.
12. R.R. Wang, M. Naphade and T. Huang. Video retrieval and relevance feedback in the Context of a post-integration model. MMSP2001. Cannes. 2001.
13. J.S. Wachman and R.W. Picard. *Tools for browsing a TV situation comedy based on content specific attributes.* Multimedia Tools and Applications. 13. 3. 2001.