To Seek, Perchance to Fail: Expressions of User Needs in Internet Video Search

Christoph Kofler, Martha Larson, and Alan Hanjalic

Multimedia Information Retrieval Lab, Delft University of Technology,
Mekelweg 4, 2628 CD Delft, The Netherlands
{c.kofler,m.a.larson,a.hanjalic}@tudelft.nl

Abstract. This work investigates user expressions of content needs in Internet video search, focusing on cases in which users have failed to meet their search goals, although relevant content is reasonably certain to exist. We study expressions of user needs in the form of requests (i.e., questions) formulated in natural language and published to Yahoo! Answers. Experiments show that classifiers can distinguish requests associated with search-goal failure. We identify a group of 'easy-to-predict' requests (cases for which the classifier predicts search-goal failure well) and compile an inventory of strategies used by users to express search goals in these cases. In a final set of experiments, we demonstrate the feasibility of predicting search-goal failure based on query-like representations of the original natural-language requests. The results of our study are intended to inform the future development of indexing and retrieval techniques for Internet video that target difficult queries.

Keywords: Multimedia retrieval, Internet video, user information need, search-goal failure, crowdsourcing.

1 Introduction

A better understanding of the ways in which users express their needs for Internet video content can inform the development of more effective video indexing and retrieval algorithms. In this work, we investigate how users have formulated their needs in cases in which they have failed to meet their search goals, despite the fact that relevant content is reasonably certain to exist on the Internet. In particular, we focus on expressions of content needs for which it is easy to automatically predict search-goal failure using a conventional text classifier. We are interested in obtaining a better understanding of the strategies that users deploy to express search goals in failure-prone cases, with an eye to our ultimate aim of developing retrieval techniques for Internet video (e.g., query prediction) that will address such cases specifically.

Our collection of expressions of user needs for Internet video is a set of requests for help finding videos that have been posted to the popular question-answering forum Yahoo! Answers (http://answers.yahoo.com/). Alternatively, expressions of needs could have been collected via a user study or transaction log analysis.

P. Clough et al. (Eds.): ECIR 2011, LNCS 6611, pp. 611–616, 2011.

Yahoo! Answers, however, provides us with a good trade off between data set size and resources invested. Critically, our data set gives us direct access to information concerning failure of users' previous attempts to meet their search goals, which would not be available in a transaction log. We use this data set to investigate three research questions: *Q1: Is search-goal failure predictable given user-expressed content needs?*, *Q2: How are content needs expressed in failure-prone cases?*, *Q3: Is it possible to predict search-goal failure given a conventional keyword query rather than a natural-language expression of a content need?*

Work related to our investigation includes transaction log analysis for multimedia retrieval [1, 2] and especially work dealing with video-search tactics [5] and image-search failure [4]. Our work is set apart from previous studies because of its use of user needs that are expressed in natural language and are also explicitly associated with search-goal failure. We adopt our basic definition of user needs and goals from [3], which defines a search goal as, 'an atomic information need, resulting in one or more queries'.

The contributions of this paper are threefold. First, we introduce, for the first time to the best of our knowledge, the use of an Internet question-answering forum for studying expressions of user content needs in the area of video search. Second, we demonstrate experimentally that it is feasible to predict search-goal failures both on the basis of a content need (expressed in natural language) and a query (expressed as a keyword set). Third, we formulate a qualitative characterization of strategies that users use to express their content needs by carrying out a manual analysis of requests associated with 'easy-to-predict' search-goal failure. The paper is structured as follows. We describe our experimental framework in the next section, then we present and discuss our experiments and the results of our analysis and, finally, we finish with conclusion and outlook.

2 Data Set and Experimental Set-Up

To create our data set, we used simple heuristics to collect requests from Yahoo! Answers that were identified as related to video search. Requests were considered video-search related if they contained the words 'find' and 'video', but not words like 'game' or 'camera'. The resulting data set is not an exhaustive set of all video-search requests on Yahoo! Answers (our heuristics are not perfect), but does provide us with a sizeable sample, which we take to be representative. Within this set, we are particularly interested in requests in which the user makes a statement that reveals that previous search efforts have not succeeded, as in this example, 'Does anyone know where I can find a video that has the monkey from an old pizza commercial in it? I think he was a sock monkey and I'm not sure what pizza company he was from [...]. I can't find the video anywhere.' In this request, we consider, 'I can't find the video anywhere.' to be a *failure statement* concerning the user's search goal. We refer to requests containing such failure statements as search-goal-failure requests, designated `+sgf` requests. In contrast, the request, 'Where do i find the video of bon qui qui at king burger? which website ?' does not contain a search-goal failure statement and is designated a `-sgf` request. It could be argued that if users generally try to find the video

themselves before posting a request on a forum, then all requests on Yahoo! Answers reflect, in a sense, failed search. However, we avoid assumptions about users' search histories, but rather take at face value what they state in their requests. We interpret the presence of a failure statement in a request to signal a particularly confounding case that can be considered 'search-goal failure prone' and deserving of special attention within our study.

We identified `+sgf/-sgf` requests in our data set using the crowdsourcing platform Amazon Mechanical Turk (http://www.mturk.com/). Workers were first recruited with a trial run that served to ensure that they understood the concept of search-goal failure. Each request was annotated by three recruited workers and we adopted the majority opinion. The final data set contained a total of 592 video search requests, with 213 labeled as `+sgf` and 379 as `-sgf`. Note that `-sgf` is not equated with a satisfied search goal, but rather with an unknown status. For each request we also asked our recruited workers to create a short keyword query that could help to find the content the requester is looking for. On average, term overlap between worker-suggested queries was approximately 63%. We merged the three queries to create a pseudo-query that allows us to carry out experiments on encodings of the user need that are more query-like than the original request, which is expressed in natural language and often quite verbose.

For each request, our data set includes four representations: (1) `need_exp_orig` the original expression of the user information need (i.e., video-search request from Yahoo! Answers) (2) `need_exp_edit` a version of `need_exp_orig` that has been normalized by manually removing any failure statement (3) `query_orig` the pseudo-query and (4) `query_exp` an expanded version of the pseudo query. Expansion was carried out using the top-ten videos returned when YouTube (http://www.youtube.com/) was queried with the pseudo-query. If, within the top-ten, a term in a title or tag list of a YouTube result had a higher-than-average co-occurrence with a query word, it was used to expand the query. How and why we use these representations is further elucidated in the next section.

Our experiments involve `+sgf/-sgf` classification. We use feature vectors of term frequencies, applying feature selection but no other preprocessing. We experiment with two classifiers, standard for text classification problems, a decision tree (J48) and a support vector machine (SVM), using Weka (http://www.cs.waikato.ac.nz/ml/weka/) as our implementation. For feature selection we use Weka's `cfssubset`, applying subset selection as the feature evaluator and best first search to select features. The decision tree gives us information about which terms are indicative and the SVM represents a state-of-the-art classifier for text. Results are generated using 5-fold cross-validation.

3 Results and Discussion

The results of our experiments, reported in terms of accuracy, are summarized in Table 1. We compare each condition with a naïve baseline under which all items are classified into the dominant class, `-sgf`. The naïve baseline achieved an accuracy of 64% (379 `-sgf` requests/592 total requests). All improvements

over the random baseline are statistically significant with respect to Wilcoxon signed-rank test ($p := 0.05$). The first two lines of Table 1 address research question Q1. We see that a standard classifier has the ability to distinguish between `+sgf` and `-sgf` user needs as expressed by original Yahoo! Answers requests (cf. `need_exp_orig` in Table 1). Indicative terms included 'can't', 'cannot' and 'find', suggesting that the classifier is exploiting features from failure statements to identify `+sgf` requests. In order to focus on specifically the expression of content needs, we removed failure statements from the requests by hand and ran the classification experiment again (cf. `need_exp_edit` in Table 1). The classification performance deteriorated sharply, but we can still answer Q1 with the observation that it appears feasible to automatically generate a prediction of failure using user expressions of Internet video needs.

Table 1. Classification accuracy, true positives (TP), true negatives (TN)

	J48 (TP/TN)	SVM (TP/TN)
`need_exp_orig`	82.8% (168/322)	85.3% (165/340)
`need_exp_edit`	65.2% (45/341)	73.3% (67/367)
`query_orig`	64.0% (0/379)	68.8% (28/379)
`query_ext`	65.0% (24/361)	77.5% (84/375)

We further investigated 'easy-to-predict' search goal failure by carrying out an analysis of those expressions of user needs that the SVM classifies correctly. We chose this group because it will be the first focus of our future research, which will be guided by the pragmatic assumption that failure needs to be diagnosable before it can be explicitly addressed. We analyzed 67 true positives by hand, examining the original answers on Yahoo! Answers and also using general Internet search to eliminate examples for which it is highly plausible that search failure can be attributed to the non-existence of relevant material, rather than the particular expression of the user need in the request. What remained was a set of 40 requests for which we are reasonably certain that relevant content exists. We coded the requests with a set of 'expression strategy' labels that described the ways that users express their content needs in the requests. We iteratively passed through the examples, refined the coding until all examples were covered by at least one label and no two labels could be reasonably conflated. The result was an inventory of strategies that provide an answer to research question Q2. The strategies are listed in Table 2 and typical examples (which has been edited slightly for length) are given for each. We also note whether requests using that strategy appear to target known items and/or ad hoc sets of videos. It is important to give this inventory the appropriate interpretation. It is a list of expression strategies that we observed are associated with search goal failure. The inventory is not exhaustive and the strategies are not necessarily mutually exclusive. We do not claim that the strategies are either necessary or sufficient to diagnose search-goal failure. Instead, this inventory provides assurance that if we create retrieval algorithms that address user needs expressed with these

Table 2. Strategies used by users in the expression of Internet video search needs

Strategies	**Example requests from Yahoo! Answers**
Uses production information (e.g., title words)	Can someone lhelp me find the official music video to this James Blunt song? its Goodbye My lover. (*known item*)
Includes identities (i.e., names of people appearing in or mentioned in the video)	Where can I find the actual video of Taylor Swift and Kanye West? The VMA awards a few weeks ago when Kanye interrupted Taylor and said that Beyonce had a great music video too... (*known item*)
Describes what is depicted visually in the video content	I'm looking for a video of a big robot walking through a mall. Can't find it anywhere, thanks for any help. It was a demo of this robot walking through a mall. It was pretty big, about 8ft tall making it's way through a mall. (*known item*)
	Where can I find a video of a hen protecting her chicks? I'm looking for a video of a hen protecting her chicks under her wings... (*ad hoc*)
Specifies theme or topic	i watched a featured video on youtube yesterday about a future shelter that could be built in case of any disasters. it could house up to 4000 people... (*known item*)
Specifies particular aspect of a specific version of a video	Anyone know where I can find the video of Spongebob and Patrick making fun of Texas? Just the parts where they're making all the Texas jokes, not the whole video. Everything I find has been dubbed over! (*known item*)
Includes a quote in the description	Does anyone know where I can find the video of Jack Buck's post 9/11 speech? I found the video of his poem but it does not include the famous "Should we be here?" quote. (*known item*)
Specifies a skill that the user wishes to acquire	Where can I find a diagram/video/etc. that shows the parts of a motorcycle? I'm about to start learning to ride a motorcycle, and I'd like to know where things are on one! (*ad hoc*)

strategies that we will indeed be working on at least some difficult cases of video search, rather than exclusively on cases for which existing algorithms already achieve good results.

In a final set of experiments, we investigate a separate, but related issue. We examine the feasibility of predicting search failure using expressions of needs closer to what users generally supply to search engines (i.e., conventional keyword queries). To this end, we perform classification experiments on the pseudo-queries corresponding to the requests (cf. `query_orig` in Table 1). Here, the performance drops off dramatically. We conjecture that the queries are simply too short to provide the classifier with enough material to generalize and we carry out an additional experiment using the expanded pseudo-queries (cf. `query_exp` in Table 1). The performance still leaves much room for improvement, but does serve to provide a positive answer to research question Q3. The results suggest that it is possible to automatically generate a prediction of search failure given an expression of the user need of the size and form of a conventional query.

4 Conclusion and Outlook

We have presented an investigation of the relationship between user expression of Internet video needs and search-goal failure, for which we used an Internet question-answering forum as a novel source of a collection of expression of user information needs. We have determined that it is feasible to automatically generate a prediction of search failure for user needs, both when they are expressed as natural-language video requests and when they are encoded as conventional keyword-based queries. A manual analysis of cases in which users failed to achieve their search goals despite the existence of relevant material on the Internet yielded an inventory of strategies. This inventory provides a picture of how users express their Internet video needs that emphasizes the perspective of difficult cases. We do not claim that it is possible to uniquely identify the exact source of search failure for any given user need or query. For example, the word 'series' turned out to be indicative of search failure for the expanded pseudo-queries (cf. `query_exp` in Table 1). This may reflect either that 'series' are in general difficult for users to find on the Internet, or that the strategy of specifying a series title or description (possibly in lieu of detailed episode information) is not particularly effective with search engines currently available on the Internet. The main conclusions of our study are that failure-prone Internet video search is potentially predictable and that users formulate information needs making use of general strategies that could be targeted by retrieval algorithms in order to improve search performance on difficult Internet video queries. Future work will involve both the extension of the analysis to a larger data set and also more detailed consideration of the answers to the requests posted on Yahoo! Answers. The results of our investigation will serve to inform long-term work on the development of video retrieval techniques for difficult queries.

Acknowledgments. The research leading to these results has received funding from the European Commission's 7th Framework Programme (FP7) under grant agreement no. 216444 (EU PetaMedia Network of Excellence).

References

1. Huurnink, B., Hollink, L., van den Heuvel, W., de Rijke, M.: Search behavior of media professionals at an audiovisual archive: A transaction log analysis. Journal of the American Society for Information Science and Technology 61(6), 1180–1197 (2010)
2. Jansen, B.J., Goodrum, A., Spink, A.: Searching for multimedia: analysis of audio, video and image web queries. World Wide Web 3(4), 249–254 (2000)
3. Jones, R., Klinkner, K.L.: Beyond the session timeout: automatic hierarchical segmentation of search topics in query logs. In: CIKM 2008: Proceeding of the 17th ACM Conference on Information and Knowledge Management, pp. 699–708. ACM, New York (2008)
4. Pu, H.-T.: An analysis of failed queries for web image retrieval. J. Inf. Sci. 34(3), 275–289 (2008)
5. Wildemuth, B.M., Oh, J.S., Marchionini, G.: Tactics used when searching for digital video. In: IIiX 2010: Proceeding of the Third Symposium on Information Interaction in Context, pp. 255–264. ACM, New York (2010)