An Image Retrieval System Based on Explicit and Implicit Feedback on a Tablet Computer

Feiyan Hu*
Queen Mary U London School of Electronic Engineering and Computer Science
London, UK
Beijing University of Posts and Telecommunications
Beijing, P.R. China
ee08b317@bupt.edu.cn

Alan F. Smeaton
CLARITY: Centre for Sensor Web Technologies School of Computing, Dublin City University
Dublin, Ireland
alan.smeaton@dcu.ie

Yan (Cindy) Sun
Queen Mary U London School of Electronic Engineering and Computer Science
London, UK
yan.sun@eecs.qmul.ac.uk

ABSTRACT

Our research aims at developing a image retrieval system which uses relevance feedback to build a hybrid search /recommendation system for images according to users' interests. An image retrieval application running on a tablet computer gathers explicit feedback through the touchscreen but also uses multiple sensing technologies to gather implicit feedback such as emotion and action. A recommendation mechanism driven by collaborative filtering is implemented to verify our interaction design.

CCS CONCEPTS

• **General and reference** → **Design**; **Experimentation**; **Measurement**; • **Human-centered computing** → **Human computer interaction (HCI)**.

KEYWORDS

Tablet computer, sensor, face detection, recommendation

ACM Reference Format:
Feiyan Hu, Alan F. Smeaton, and Yan (Cindy) Sun. 2021. An Image Retrieval System Based on Explicit and Implicit Feedback on a Tablet Computer. In *Proceedings of iHCI 2012: 6th Annual Irish Human Computer Interaction (HCI) Conference (iHCI 2012).* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Much research has been carried out examining the requirements of people when we are seeking useful information, but information retrieval and search is still a complicated and multi-faceted activity because it is personalized and strongly depends on the preference and needs of the individual. The search engine has become the main tool for people to use when seeking any kind of information, from the rich resources of the Internet to closed digital libraries. In the search process, the user sees a ranked list of results, ranked according to similarity to the query words that the user typed into the search box. Recommendation can be defined as finding items that a user might like based on how the system understand the user's interests and can be re- garded as a development of the search engine. Nowadays, the biggest challenge for both recommendation and search systems is that the engine cannot take the user's unique character or the user's current contexts into account, in order to get a more personalized recommendation or search result. There have been many studies which have examined the characteristics of user queries to search engines on the web and most of these have found that queries are short and do not elaborate the full information need of the user [4] and the user may have a completely different intention for the set of words chosen in the same query [3]. For example, the query "windows" may search for car windows or from a software developer might refer to the operating system Windows. With the help of a personalized profile, the quality of search results would be greatly improved.

One of the other drawbacks with current search and recommendation systems is that it is difficult for a system to assess the results of a recommendation or search listing if at all, since relevance feedback is not generally incorporated into search systems. In relevance feedback the user feeds back information on the usefulness of items presented so that the system can use these judgments in order to refine its search ranking or recommendations. However, when relevance feedback is used it is almost invariably a binary response – a user "likes" something or does not, or the user can indirectly imply a positive judgment on an item by sharing their rating of an item with others. This leads to a concern about such computer based recommendation for fear of the so-called "filter bubble", whereby all liked items are treated equally rather than some being more liked than others and the resulting search and recommendation algorithms cannot exploit degrees or relevance among presented items. Ideally it would be preferable to allow a user to feed back an indication of such degrees of relevance or even to feedback which facets or a recommended item make it relevance but in a search task where the cognitive load on a user is high and there are many conflicting stresses, the choice is to minimise the onus of user feedback by having relevance judgments as a binary feedback, if at all.

*This work was carried out while F.H. was a student at CLARITY: Centre for Sensor Web Technologies, Dublin City University

iHCI 2012, June 20–21, 2012, Galway, Ireland

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

In this paper we describe a system for obtaining degrees of relevance form a user in an image search task by capturing implicit as well as explicit relevance feedback and combining these within the search/recommendation algorithm to choose items to present to a user. We now describe some of the relevant background to our work.

2 BACKGROUND

A key technology to improve the ability of an information retrieval or a recommend system is query expansion, which uses the original query from the user, and adds relevant words which can produce a more focuses query in order to better describe the implicit semantics from the original query. Typically this is achieved by the user judging sample items and inferring an expansion of the query based on term distribution and so relevance feedback is a vital technique which can be applied to expand query [6] and this has been known for a long time. The idea is to adapt the system to the specific user preferences making more important weights or features that reflect the actual user needs in order to achieve higher precision. In other words, relevance feedback depends on the individual interaction between user and system to improve the performance of the system by collecting judgments by user on the returned results from using original query. The basic premise for relevance feedback is “You may not know what you’re looking for, but you will know when you see it”. According to the different ways, in which users interact with retrieval system, feedback can be classified into explicit, implicit and blind.

Explicit feedback is obtained from users indicating the relevance of a document retrieved for a query. This type of feedback is defined as explicit only when the assessors know that the feedback provided is interpreted as relevance judgments. The most vital concept of explicit feedback is that assessors evaluate the quality of the retrieved information using predefined criteria and usually a binary or graded relevance system is used. Binary relevance uses options such as ‘like’ or ‘dislike’ to imply whether the retrieved information is relevant or not. Graded relevance uses more option instead of two to show the relevance of the retrieved system using human-like language, numbers or letters to represent the level of relevance. Google has launched its graded explicit feedback service called searchWiKi (now called Google Stars), which takes the form of letting assessors place documents from the result list in order of relevance.

Implicit feedback is related to the behavior, action or reaction of the user, either visible or invisible or both. For example if an assessor clicks a certain document, this implies their interest in the clicked document. If the duration of viewing some documents is longer than others, this means that he might be more interested in the content. The key differences between implicit and explicit relevance feedback include:

- The user is not assessing relevance for the benefit of the IR system, but only satisfying their own needs, and
- The user is not necessarily informed that their behavior will be used as relevance feedback.

Explicit relevance feedback can be seen as a conscious reaction from the assessor, while implicit feedback is a response of the sub-conscious but both should be used to gather feedback, where possible.

One popular form of information retrieval and recommendation is image retrieval. An image retrieval system is a system for browsing, searching and retrieving images from a large database of digital images. We can classify image retrieval into text-based retrieval and content-based image retrieval. In text-based retrieval the query is always text, and search matches the query with the meta data of the image such as the title, user tags or comments in the case of the Flickr or Panoramio systems.

Content-based image retrieval (CBIR) uses the content of the image (texture, shape, or color) instead of textual description to identify the similarity between query and image. For example a user can draw a picture and ask CBIR to find something similar or use an image as a query to find similar ones. CBIR doesn’t use the meta data like keywords, tags or other descriptions associated with the image, which means CBIR tries to ‘understand’ the content within a image.

Relevance feedback can also be used in CBIR where a user can inform the system about the relevance of given images and the remaining unseen images re-ranked. Managing a user’s navigation through an image database whereby the user marks some images as relevant resulting in a re-ranking of the unseen allows the user to follow a scent of an information need and then backtrack to an earlier point in their search and spawn off an information need in another direction. This is referred to as ostensive relevance feedback, an iterative process allowing multiple facets of a user’s information need to be explored in sequence [1].

Current CBIR always uses color, texture and shape, which are all low-level semantics, to calculate the image distance from the query although recent research focuses on how to describe using high-level semantics.

3 SYSTEM ARCHITECTURE

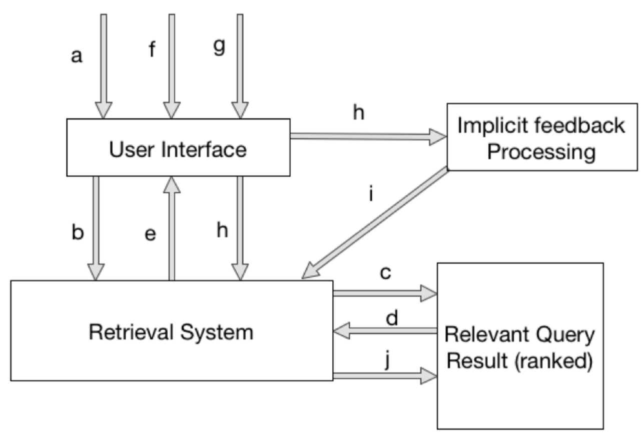

Figure 1: System architecture

We built an image retrieval system as an iPad-2 application combining explicit and implicit user feedback. The system is divided into 4 parts as shown in Figure 1. The user interface (A) is responsible for user input and detects user implicit feedback. The retrieval system (B) implements the retrieval algorithm by adjusting the query vector. Relevance calculation (C) is completed as a component associated with the database. The implicit data processing component (D) takes charge of the computation of user implicit information detected on the user interface from on-board sensors.

The following shows the operation of the system. (a) User sends the query to search for an image (assuming the query is text). (b) Interface sends the query to the retrieval system. (c) Because it is the original query, the retrieval system forwards the query to the component to calculate relevance. (d) The component cooperates with the database returning a ranked relevant result. (e) The retrieval system looks up the indexed database finds the picture and returns it to the User Interface. (f) and (g) User Interface accept explicit and implicit feedback and (h) sends the explicit feedback to the retrieval system and at the same time sends the implicit feedback to the data processing component. (i) Processing component sends the processed data to retrieval system. (j) Retrieval system processes the original implicit feedback and adjustment algorithm and obtains a new query vector, then sends it to the last component to re-compute similarity.

Because the time to process implicit feedback should be considered, this means there is a time difference between arrival of explicit and implicit feedback so a threshold should be set to identify when to use the implicit feedback as a degree reference. According to the well-known Rocchio's algorithm [2], the optimized query vectors we want to find should maximize similarity with relevant images while minimizing similarity with non-relevant images. More formally, this can be stated as $Q_{opt} = \arg\max[sim(q, C_r) - sim(q, C_{nr})]$. After introducing a weight calculated from implicit feedback to every image, below is the implicit feedback value calculate by our algorithm based on that presented by Manning et al. in [5]:

$$\vec{Q_m} = \left(a * \vec{Q_o}\right) + \left(b * \frac{1}{|D_r|} * \sum_{\vec{D_j} \in D_r} w_j \vec{D_j}\right) - \left(c * \frac{1}{|D_{nr}|} * \sum_{\vec{D_k} \in D_{nr}} \vec{D_k}\right) \tag{1}$$

where $w_j = \frac{IMP_j}{max[IMP]}$ and IMP are the implicit feedback value calculate by the algorithm. Most retrieval systems set $c < b$ but for most image retrieval systems and indeed in our own work, negative feedback is filtered and only positive feedback is taken into consideration, as we do here.

In its implementation on the iPad-2 platform, the application is separated into 3 components, which are responsible for (1) face detection, (2) built-in sensor detection and (3) capturing user interaction. Face detection embeds high-performance, real-time face detection on each frame from the front-facing video camera on the IPad-2 using the OpenCV library with a new frame about every 40ms, which is satisfactory. Video flows are optimized by grey-scale and resizing of each frame before detection, in order to improve performance and save storage space on the tablet. After a face detected, the frame is saved to the device and all related sensor data and the captured face image are uploaded for further face analysis. The facial analysis can be achieved by online detection, and a set of face information such as the positions of eyes, mood and smile are logged into the database.

The iPad-2's built-in sensors including the accelerometer, is used to detect movement of the user to help decide whether the user is looking at the device. When the user want to interact with the system, there must exist a activity to trigger an event, for example touching a screen or clicking a buttons. Based on the data we gathered from the iPad-2, we call a collaborative filtering recommendation system to test whether the users' interest which we capture, helps with retrieval.

4 EXPERIMENT DESIGN

We ran an experiment where 11 users were asked to use the software on the iPad-2 to browse some pre-stored images with no overall task, just to browse images at their own pace. The total number of images is about 1200, classified according to content. Test users were informed the functionality of the software, including how to change the image to the next one and using gesture to manipulate the image. The software interface includes an image presentation area, a button to move to the next image, and three indicators to show the running situation of the program. The image presentation layer is interactive within which the user can zoom in, zoom out or drag the image using their fingers. The three indicators are click-count indicator, which used to show how many images that the user has browsed, holding-statue indicator which shows whether the user is holding the device or not, and a face detection indicator which displays whether the front camera detects a face in order to help the user to change position to be detected easier.

Each user was asked to browse 500-600 images, randomly extracted from the folders and approximately half of the whole image collection. Figure 2 shows the distribution of interests for each classification across our 11 users, where darkness of the color implies higher interest. From this picture we can identify what kind of picture our users are interested in. Users 1, 5 and 10 have overall stronger interest in the images than the others while the leftmost column is the aggregation of all users on a per-category basis and shows our 11 users are more interested in car races, chocolate and wine than autumn, dog or insect. While this initial experiment does not tell us much except what kinds of pictures a particular group of people prefer, it shows that we can capture and combine both explicit, and implicit relevance feedback. Based on this we moved on to build an image recommendation list for each user. Our system uses the iPad-2 sensors to detect whether the user is holding the iPad or not and if so whether the user is looking at the picture being shown on the screen and what the user's facial expression is. A real-time API call to face.com gives us an indication of whether a face is detected, how far it is from the screen/camera, and whether eyes, ears, tip of nose etc. are visible, and from this and the accelerometer readings we can infer whether the user is holding the iPad and looking at the screen. If so then face.com also returns the gender, estimated age and mood categorisation of the face. Figure 3 shows a sample image from the iPad camera[1], consisting of a face with various parts of the face (eyes, nose) indicated by red dots.

[1]This image is deliberately blurred for anonymous paper submission

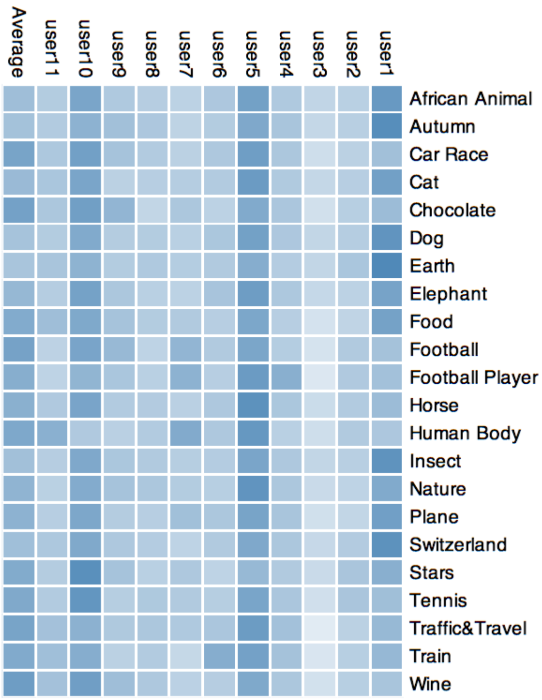

Figure 2: Interests of users per category.

Figure 3: Sample image from iPad-2 front-facing camera. The Face.com API categorises this as a 49-year-old male who is angry, and is not smiling. Note red dots indicating centre of eyes, nose tip and edges of lips.

5 CONCLUSIONS AND FUTURE WORK

In this paper we have designed and implemented an image retrieval system where a combination of both explicit and implicit relevance feedback are used to refine and re-rank the as yet unseen images in a user's search. The user interface component runs on an iPad-2 with a front-facing camera and is used to sense explicit feedback via a "likes"-type button on the screen which the user taps, and to sense implicit feedback based on a combination of how long the user holds the iPad and looks at the image and a categorisation of the user's facial expression while viewing. This is achieved via an API call to face.com where results are returned in JSON format in real time. The novelty in this work is the combination of explicit and implicit feedback and in fact the user could search the image collection and move from image to image in the ranked list of outputs just based on facial expression.

There is much further work that needs to be done before we can claim that we have balanced and combined implicit and explicit user feedback in an operational setting. The balance between explicit and implicit feedback is currently a 50:50 ratio but ultimately this balance will depend on the user, characteristics of the search (broad or narrow topic, urgent or relaxed task requirement, exhaustive or "first-willdo" search requirement, etc.), the specificity of the search (i.e. the absolute vs. the relative similarities of images to the query) and the stage of the search, i.e. how far into the actual search the process has proceeded. Even within the broad term of "implicit relevance feedback" we need to learn how to fuse the sensor inputs from holding the iPad and the outputs from the face analysis on http://face.com. We also need to cater for ostensive relevance feedback, referred to earlier as allowing multiple facets of a user's in- formation need to be explored in sequence [1]. A natural phenomenon in information seeking is that as we browse the output of a search or recommendation process, our information needs shift and evolve as we become more informed. At some point we may wish to pursue more than one aspect or facet of our search results so ideally we should be able to bookmark the present point in our search, proceed to explore one aspect and then return to the bookmark to pursue a different aspect of the search. For example when doing a web search for a recipe for baking cookies, you see a link to a page about the pros and cons of using a gas oven vs. an electric and you make a mental note (a search bookmark) to return to that link because you are considering changing your cooker.

Our plan in the future is to carry out user studies but the challenge is in replicating a real user image search task on a benchmark dataset that is publicly available.

ACKNOWLEDGMENTS

This work is supported by Science Foundation Ireland under grant 07/CE/I1147.

REFERENCES

[1] Paul Browne and Alan F Smeaton. 2004. Video information retrieval using objects and ostensive relevance feedback. In *Proceedings of the 2004 ACM symposium on Applied computing*. 1084–1090.

[2] SF Dierk. 1972. The SMART retrieval system: Experiments in automatic document processing. *IEEE Transactions on Professional Communication* 1 (1972), 17–17.

[3] Myunggwon Hwang, Pankoo Kim, and Dongjin Choi. 2011. Information retrieval techniques to grasp user intention in pervasive computing environment. In *2011 fifth international conference on innovative mobile and internet services in ubiquitous computing*. IEEE, 186–191.

[4] Bernard J Jansen, Amanda Spink, Judy Bateman, and Tefko Saracevic. 1998. Real life information retrieval: A study of user queries on the web. In *ACM Sigir Forum*, Vol. 32. ACM New York, NY, USA, 5–17.

[5] Hinrich Schütze, Christopher D Manning, and Prabhakar Raghavan. 2008. *Introduction to information retrieval*. Vol. 39. Cambridge University Press Cambridge.

[6] Alan F Smeaton and Francis Crimmins. 1997. Relevance feedback and query expansion for searching the web: A model for searching a digital library. In *International Conference on Theory and Practice of Digital Libraries*. Springer, 99–112.