Aslib Journal of Information Management

Improving retrieval relevance using users' explicit feedback

Vimala Balakrishnan Kian Ahmadi Sri Devi Ravana

Improving retrieval relevance using users' explicit feedback

Vimala Balakrishnan, Kian Ahmadi and Sri Devi Ravana
University of Malaya, Kuala Lumpur, Malaysia

Received 8 July 2015
Revised 28 September 2015
Accepted 2 November 2015

Abstract

Purpose – The purpose of this paper is to improve users' search results relevancy by manipulating their explicit feedback.

Design/methodology/approach – CoRRe – an explicit feedback model integrating three popular feedback, namely, Comment-Rating-Referral is proposed in this study. The model is further enhanced using case-based reasoning in retrieving the top-5 results. A search engine prototype was developed using Text REtrieval Conference as the document collection, and results were evaluated at three levels (i.e. top-5, 10 and 15). A user evaluation involving 28 students was administered, focussing on 20 queries.

Findings – Both Mean Average Precision and Normalized Discounted Cumulative Gain results indicate CoRRe to have the highest retrieval precisions at all the three levels compared to the other feedback models. Furthermore, independent *t*-tests showed the precision differences to be significant. Rating was found to be the most popular technique among the participants, producing the best precision compared to referral and comments.

Research limitations/implications – The findings suggest that search retrieval relevance can be significantly improved when users' explicit feedback are integrated, therefore web-based systems should find ways to manipulate users' feedback to provide better recommendations or search results to the users.

Originality/value – The study is novel in the sense that users' comment, rating and referral were taken into consideration to improve their overall search experience.

Keywords Rating, Referral, Case-based reasoning, Comment, Explicit feedback, Retrieval relevance

Paper type Research paper

Introduction

With the advent of the internet, finding relevant documents or items among those provided by search engines can be a daunting task. A simple example would include a researcher performing a literature search using the library database containing hundreds or even thousands of documents. Various studies have attempted to ease the task of information searching by finding ways to improve the search results, particularly by exploiting users' feedback. In text or information retrieval, users' feedback is generally categorized as implicit and explicit. Implicit feedback such as click-data (i.e. a series of pages selected for viewing in a search session) (Balakrishnan and Zhang, 2014; Buscher *et al.*, 2012; Bidoki *et al.*, 2010; Jung *et al.*, 2007) and scrolling (Buscher *et al.*, 2012; Guo and Agichtein, 2012) have been used to retrieve, filter and recommend a variety of items such as web documents, movies and books, among others. Implicit feedback can be collected unobtrusively, however it contains a huge amount of noise which may affect results or documents relevancy (Jung *et al.*, 2007).

On the other hand, explicit feedback makes a prediction based on users' judgments using ratings, or rankings, for example. Explicit feedback is deemed to be more accurate than implicit feedback, however it requires users to perform additional activities other than viewing or reading a document. Studies focussing on users' explicit feedback are very few, with many investigating the correlations between

Aslib Journal of Information Management
Vol. 68 No. 1, 2016
pp. 76-98

2050-3806
DOI 10.1108/AJIM-07-2015-0106

implicit and explicit feedback (Claypool *et al.*, 2001; Fox *et al.*, 2005; Lagun *et al.*, 2013; Xu *et al.*, 2010; Lagun and Agichtein, 2011; Guo and Agichtein, 2012; Buscher *et al.*, 2012; Jawaheer *et al.*, 2014). Furthermore, most of these studies focussed on a single feedback approach, for example, Claypool *et al.* (2001) analyzed users' ratings. The review of literature shows that an integrated model (i.e. more than a single implicit or explicit feedback) and hybrid models (i.e. combination of implicit and explicit feedback) can improve the retrieval relevance for a user search (Balakrishnan and Zhang, 2014; Buscher *et al.*, 2012; Lagun *et al.*, 2013).

Case-based reasoning (CBR) is a highly intelligent technique for problem-solving using old similar experiences to address new problems. The technique suggests that current problems can be assessed by finding previous cases relevant to the current problem, and leverage that case to inform a solution, and update the memory as one learns from the experience (Aamodt and Plaza, 1996; Yan *et al.*, 2014). CBR has been widely and successfully used in various fields such as medicine (Henriet *et al.*, 2014; Bixby, 2013), knowledge-based system (Lupiani *et al.*, 2014) and information seeking behavioral studies (He *et al.*, 2008; Belkin *et al.*, 1995). CBR is known to quickly propose an effective new solution, hence the process and time of decision making are greatly reduced (Kolodner, 1999), however very few studies have looked into using CBR in information retrieval (Daniels and Rissland, 1995; Beebe *et al.*, 2011).

The current study was undertaken to improve users' search results retrieval relevance by developing an integrated feedback model, aptly named Comment-Rating-Referral (CoRRe). CoRRe focusses on three types of explicit feedback, described as follows:

(1) Comment – textual comments provided by users regarding an item/service. Comments are commonly seen at popular sites such as Amazon.com, eBay, YouTube, TripAdvisor, etc.

(2) Rating – user ratings provided based on specific scales. Ratings are popularly used as they are often easily quantifiable. For instance, Amazon.com uses a five-point scale whereas Facebook uses the "Like" system.

(3) Referral – the concept of referral refers to the users' recommendations of a product/service. Generally, positive recommendations are often provided when users are satisfied with a product/service, and thus resulting in higher chances of other users to purchase the same product/service (Anderson, 2014). For example, Google+ allows its users to recommend websites to their contacts. This concept can be adapted to information retrieval as well, as an item that is recommended (or referred) indicates users' satisfactions regarding its relevancy.

The feedbacks elicited from the users were then used to re-rank the original results by adapting the five-star algorithm. The results are stored and retrieved using CBR. CoRRe was evaluated by developing a prototype search engine using data from the Text REtrieval Conference (TREC) as the core document collection. A group of users were recruited to evaluate CoRRe, and the data gathered were analyzed and compared to assess the efficacy of the model. Evaluations were conducted at top *k*-levels, namely, top-5, 10 and 15 using the standard Mean Average Precision (MAP) and Normalized Discounted Cumulative Gain (NDCG) metrics. As will be shown later, CoRRe significantly improves the retrieval relevance compared to the baseline model (i.e. a model without any user feedback), and other feedback models at all the three levels.

The study is considered novel from the following perspectives:

- the concept of referral or recommendation was introduced to improve the retrieval relevance.
- three types of user feedback (i.e. comments, ratings and referrals) were integrated into a single feedback model to improve the retrieval relevance for a user query.
- the nature of users' comments (i.e. positive/negative) were taken into consideration to determine results relevance.
- a re-ranking algorithm adapted from the five-star algorithm was proposed to re-rank the original query results using users' feedback.

Background studies

Implicit feedback

Implicit feedback attempts to estimate relevancy based on a user's behavior without requiring any additional efforts by the user, hence it is considered to be a more practical method for data collection. However, most implicit feedback contain too much noise, and thus does not always indicate relevancy (Jung *et al.*, 2007; Bullock *et al.*, 2011). For instance, it has been argued that display or viewing time does not necessarily imply relevancy as a searcher may spend a long time looking for something relevant on a promising site, but fails to find anything relevant (Oard and Kim, 1998). Similarly, although click-data can be used to improve search results, the accuracy of the retrievals are influenced by trust bias (i.e. users clicking on the top few results as they trust the ranking quality of the search engine) and quality bias (i.e. quality of the click-data are affected as a user's clicking behavior varies for the same topic in different search engines) (Joachims *et al.*, 2005; Gao *et al.*, 2009).

Explicit feedback

Unlike implicit feedback, explicit feedback requires users to do additional work, resulting in low number of responding users (Ricci *et al.*, 2011). Moreover, this technique incurs a high cost in the sense that users frequently need to examine several documents in one search session, and thus often ignoring the requirement to provide explicit feedback (Manning *et al.*, 2009). In fact, some search engines and recommendation systems depend on the expensive explicit feedback of experts to improve the system performance (Raman *et al.*, 2012). Nevertheless, it is generally accepted that explicit feedback is more accurate than implicit feedback in modeling users' interests (Buscher *et al.*, 2012), probably due to the availability of several domain-independent, objective, well-researched and documented tools, such as Likert scales or questionnaires for capturing and analyzing users' explicit feedback. In fact, one of the major advantages of explicit feedback is that it provides good indications of users' interests and improves the search results through accurate users' relevance judgments (Baeza-Yates and Ribeiro-Neto, 1999). More importantly, explicit feedback captures both positive and negative views, and also involves active user participations (Jung *et al.*, 2007). One of the most common techniques in providing feedback is rating, which is often based on a Likert scale (Jannach *et al.*, 2011 as cited in Jawaheer *et al.*, 2014). Many systems, particularly recommender systems use the five-point scale (e.g. Amazon.com and MovieLens), whereas some use binary ratings, such as Last.fm which allows users to provide their feedback via two options (i.e. love or ban a track). Unary ratings (i.e. only positive ratings) are used by Facebook (i.e. "Like") and Twitter (i.e. "Favorite").

Despite the advantages, very few studies have focussed on explicit feedback, particularly in improving search results relevance. Instead, many investigated the relationships between implicit and explicit feedback in predicting users' interests. For instance, Núñez-Valdéz *et al.* (2012) compared the performance of implicit (e.g. reading time, number of visits and number of clicks, etc.) and explicit feedback (e.g. comments, ratings and referrals) for an e-book recommender system. Some interesting results include that the average rating for content increases with the number of comments, and users also tend to recommend an item that has high ratings. It is to note that the authors did not consider the nature of the comments, that is, whether they are positive or negative. Instead the comments were measured in terms of their frequency of occurrences. It is crucial to differentiate the nature of the comments, as positive words indicate the user to be satisfied with the item, whereas negative words indicate otherwise (Shamim *et al.*, 2014; Liu, 2012). Nevertheless, the study provided useful insights into some of the explicit feedback that can be further manipulated to improve search results relevance.

Another similar study used implicit (i.e. track play count) and explicit feedback (i.e. love or ban a track) to recommend music tracks to the users, with results showing both approaches to produce similar performances, probably due to the limited data set used in the experiment (Jawaheer *et al.*, 2010). Claypool *et al.* (2001) examined the associations between various types of implicit feedback (e.g. scroll bar clicks, mouse movement time, dwell time, etc.) and users' ratings. Their findings indicate dwell (i.e. reading or viewing) time and the amount of scrolling on a page have a positive correlation with the explicit ratings, indicating that the higher the rating, the more the users spend time and scroll the page. The rating approach is generally more popular compared to referral and comments. In fact, some researchers have looked into asking the users to re-rate the results, claiming re-rating significantly improved recommendation accuracies (Amatriain *et al.*, 2009).

Literature exists on the use of text comments (e.g. users providing reviews on Amazon.com), particularly on product reviews but many focus on recommender systems (Wei and Lu, 2013; Núñez-Valdéz *et al.*, 2012; Lu *et al.*, 2009; Siersdorfer *et al.*, 2010; Garcia Esparza *et al.*, 2012; Choo *et al.*, 2014) or in finding optimal ways in analyzing the comments to understand users' decision-making process (Shamim *et al.*, 2014; Liu, 2012; Mudambi and Schuff, 2010; Ngo-Ye and Sinha, 2012). Some studies such as Núñez-Valdéz *et al.* (2012) identified relationships between users' ratings and the number of comments whereas Choo *et al.* (2014) investigated the associations between reviews (i.e. users' opinions or feedback) and comments (i.e. users' replies to a review). Some researchers applied text comments in improving retrieval accuracy for videos with promising results (Masuda *et al.*, 2008; Kimura *et al.*, 2008; Wakamiya *et al.*, 2011). To the best of our knowledge, no studies have looked into using users' comments in improving document search results retrieval relevance.

Personal recommendations can have positive influence on people in their decision-making process. For instance, a user may opt to purchase a book based on his/her friend's recommendation. In an online scenario, the same user may also opt to purchase a book based on other users' recommendations. In fact, a recent survey on consumer usage and attitudes toward online reviews revealed that a vast majority of users trust online reviews as much as personal recommendations, suggesting the importance of users' comments and also recommendations (Anderson, 2014). The referral idea has been studied in recommender systems with results showing users to be more interested in items/services recommended by people rather than a system. For instance, Lerman (2006) studied the effect of user recommended stories as opposed to system recommended stories, and found users to be more interested in the former. Similarly, Sinha and Swearingen (2001) compared movie and book recommendations from friends with online recommender systems, and found users to prefer and trust recommendations from friends more. These studies show the importance of referrals, and as far as we are concerned, the idea of using referrals has not been used in information retrieval, particularly in improving retrieval relevance.

Generally, many researchers favor implicit feedback, as it is considerably easier to gather large volume of users' feedback unobtrusively. However, explicit data provides richer insights into users' interests (Buscher *et al.*, 2012). Studies particularly on recommender systems have shown that users' explicit feedback can be used to improve recommendations, albeit with many focussing on the most common feedback technique, that is, rating. Few recent studies have started exploring users' comments (i.e. opinion mining) (Shamim *et al.*, 2014; Liu, 2012; Mudambi and Schuff, 2010) so as to be able to better understand users' sentiments and needs, while some have looked into correlations between comments with other feedback techniques in recommendation systems (Núñez-Valdéz *et al.*, 2012; Choo *et al.*, 2014), but none have looked into using comments to improve document retrieval relevance. Similarly, user referrals have been shown to be favorable among users, both offline and online (Lerman, 2006; Sinha and Swearingen, 2001), but it has yet to be used to improve retrieval relevance. Therefore, the current study aims to fill in the gaps existing in using users' explicit feedback in improving retrieval relevance. For this purpose, an integrated model combining CoRRe was developed and tested using the widely established TREC document collection. The subsequent section describes CoRRE in detail.

Research methodology

CoRRe

Figure 1 illustrates the overall flow of CoRRe. A user basically performs a search via the search engine, and an initial set of results is presented based on the baseline model (i.e. greedy algorithm). This is assuming that the search is new, and therefore no existing results are available in the case-base. If the search query is not new, then the top-5 relevant documents for the query are fetched from the case-base, and supplemented with results using the greedy algorithm. In case of an overlap (i.e. both case base and greedy algorithm have the same results for the same query), then the documents produced by the greedy algorithm will be excluded from the results. This is because documents stored in the case-base contain users' feedback, therefore their relevance weights are higher. When a user provides feedback (comment, rating and/or referral), the system captures this feedback and re-ranks the results based on the re-ranking algorithm. The enhanced results are then presented to the user for the same query. The top-5 most relevant results are automatically saved in the case-base for this new query (or updated for a repeated query). This process is iterated every time a user provides a feedback for the same keyword or query. Therefore, when another user performs a search for the same query, the system retrieves the most relevant top-5 results from the case-base (i.e. based on the previous search) and appends them with

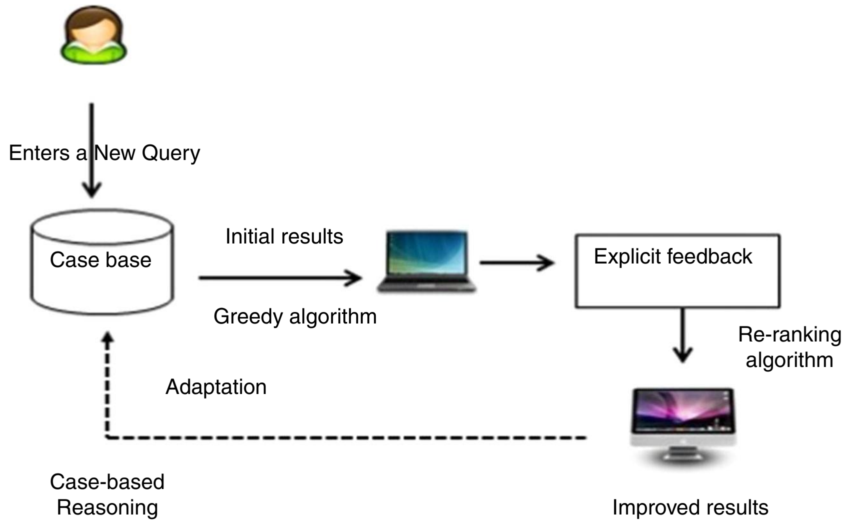

Figure 1.
The CoRRe architecture

the results of the greedy algorithm. The case base is thus capable of growing dynamically to support new queries.

It is to note that the top-5 documents are retrieved and stored based on their relevance weights (i.e. the higher the weight, the more relevant it is). For instance, assume there are only three relevant documents for Query A in the case base, then the results will contain these three documents followed by those retrieved by the greedy algorithm. The ranking of documents for Query A changes when a user provides explicit feedback, which may then result in more relevant documents being stored (i.e. added or updated) to the current case base.

Each of the main modules in Figure 1 is described in the following sub-sections.

The baseline

Most existing works in information retrieval use baseline algorithms in order to perform the initial retrieval; for example, Balakrishnan and Zhang (2014) and Kim (2014) used Term frequency-inverse documents frequency, whereas Agichtein *et al.* (2006) and Bidoki *et al.* (2010) used BM25. Others such as Lagun and Agichtein (2011) and Xu *et al.* (2010) used the Google search engine. The current study adopted the greedy algorithm as the baseline model. A greedy algorithm is one that follows the problem solving heuristic of making the locally optimal choice at each stage with the hope of finding a global optimum (further details in Dasgupta *et al.*, 2006; Chen, 2008). Though the algorithm generally does not produce an optimal solution, it may nonetheless yield locally optimal solutions that approximate a global optimal solution in a reasonable amount of time. The algorithm often works well in solving mathematical problems, however it has also been successfully used by various researchers in the field of information retrieval such as Costa (2010), Bhatt and Rusiya (2013) and Kumar and Sandeep (2013). Furthermore, the greedy algorithm was also found to be feasible to be implemented as it builds a solution one step at a time, and thus works well with the wt10g data set used in the study (i.e. wt10g is divided into different files containing data). The algorithm partitions the data, and then performs a search focussing on each partition, before retrieving the needed data and information to the user.

Users' explicit feedback

CoRRe focusses on three types of users' explicit feedback, namely, comments, ratings and referrals. These were gathered using the following scales:

Comment – A text box supporting up to 150 characters is provided for the users to comment on each document. The system checks specific keywords in a comment, and compares it against a set of sentiment words (i.e. positive and negative words). To be precise, we used the well-established semantic lexicon list containing approximately 6,800 positive and negative words by Hu and Liu (2004).

Rating – a four-way scale was used for the users to rate the relevancy of a document, that is Not Relevant (0), Low Relevant (1), Relevant (2) and Very Relevant (3), with numbers in the brackets denoting their weights (i.e. the higher the relevancy, the higher the weight).

Referral – a dichotomous scale (i.e. Yes = 1; No = 0) was used for the users to recommend a document based on its relevancy.

As in most existing systems, this feedback is optional. In the case of no feedback, results are produced by the greedy algorithm. In other words, the original results will not be re-ranked.

The re-ranking algorithm

The five-star rating used by popular sites such as Amazon.com was adapted to re-rank the original results using users' explicit feedback. To be specific, the following equation was devised to rate and re-rank the results:

$$Score = \sum_{i=0}^{n} [Rating + Referral \pm 1.5(Comment)]_{i/n} \tag{1}$$

whereby:

- Both rating and referral were given equal weights, that is, a "1" with the assumption that these two types of feedback would be more popular among the users. Moreover, the feedback was measured using fixed objective scales hence users' thoughts may not be accurately captured as opposed to comments.
- A higher weight was assigned for comment since it contains richer details pertaining to the relevancy. Additionally, comments also require more effort and time, therefore we predict users may not actively provide comments compared to rating and referral. Both positive and negative weights were used to differentiate the comments. For example, when a comment exists for a document, a value of "1" is assigned to the variable Comment. Then the system checks the comment against the semantic lexicon list (Hu and Liu, 2004) and counts the total number of positive and negative words. Comments with more positive words are then assigned a weight of +1.5, and the reverse is true for a negative comment. If the number of positive and negative words is equal, the comment is assumed to be positive.
- n: total number of users' feedback.
- i: each document.
- Score: The weighted average score for a particular query.

The scores for each of the queries and the documents are calculated and averaged into a single value, which is then depicted using the star symbols. By way of an example, assume a user rates document A as follows:

$$Rating = Relevant; Referral = Yes; Comment = Yes$$

Also assume the comments are positive, hence the score for Document A is as given below:

$$\begin{array}{ccc} \text{Relevant} = 2 & \text{Yes} = 1 & \text{Yes} = 1 \\ \uparrow & \uparrow & \uparrow \\ \multicolumn{3}{c}{[(1 \times 2) + (1 \times 1) + (1.5 \times 1)] = 4.5} \\ \downarrow & \downarrow & \downarrow \\ \text{Weight} & \text{Weight} & \text{Positive weight} \end{array}$$

The rating for document A (i.e. 4.5) based on a single user will then be indicated using the stars. If ten users rate document A, then the final score will be averaged to produce a single value.

CBR

The CBR identifies the current problem (i.e. a query in our case), finds a past case similar to the new one, and uses it to suggest a solution to the current problem, evaluate the proposed solution and update the system by learning from this experience (Aamodt and Plaza, 1996; Yan *et al.*, 2014). In the current study, the concept of CBR is adapted so that only the top-5 relevant results for each query are stored in the case base, and thus enhancing the efficiency of the retrieval process. A user's keyword-based query is represented as a series of query terms, which are mapped to their respective relevant documents (i.e. cases), if any. This was done to ensure the system responds to the users with lists of highly relevant documents based on their queries, in a more efficient manner. The relevant cases are updated whenever there is a change in the document weight (i.e. due to users' feedback). A particular case is not updated if the user provides no feedback, however it may be removed from the case base or re-ranked if other cases have been updated. It is to note that CBR was used to store highly relevant results, retrieve those results, solve a user's query search and update the results upon receiving users' feedback.

Figure 2 shows a simple illustration of what takes place in the case base. User A performs a search for Query 1, and the system displays results using the greedy algorithm (assuming a first-time search for the query). Upon providing feedback, three relevant documents are stored in the case-base. He then performs a different search for Query 2, and the relevant results are stored. The documents are ranked and stored based on the specific queries. When User B performs a similar search for Query 1, the system retrieves the top results from the CBR (i.e. solve), and appends the list with results produced by the greedy algorithm (i.e. the top-3 documents are considered highly relevant based on other users' ranking). User B then provides negative feedback for Document B, resulting in a reduction in its weight. These changes are then reflected in the case-base by re-sorting (i.e. update) the list, as indicated in the Figure. It is to note that the system is not customized for specific

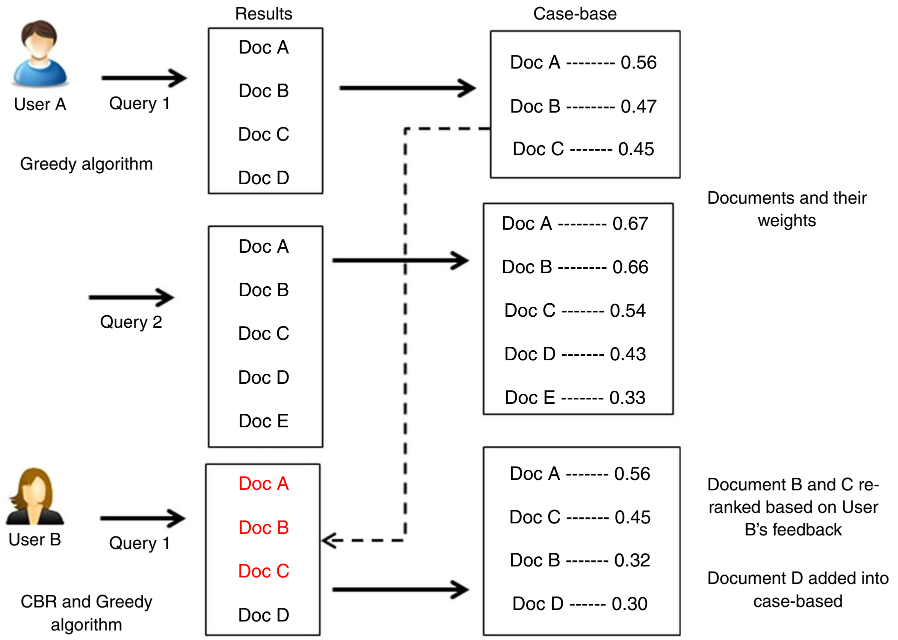

Figure 2.
An overview of the re-ranking and storage processes

users; instead it works by retrieving highly relevant documents for each query based on their weights. Therefore, when User A performs a search for Query 1 again, the most recent updated documents will be presented (i.e. Doc A, C, B, D).

Test collection

A test collection is basically an abstraction of an operational retrieval environment that enables researchers to compare different retrieval techniques in a laboratory setting. The TREC is an on-going series of workshops focussing on a list of different information retrieval research areas (i.e. tracks), and widely used to evaluate and compare the effectiveness of the retrieval systems. TREC consists of three parts: a set of documents, a set of information needs called topics and relevance judgments (i.e. an indication of which documents are considered relevant to a particular topic) (Voorhees and Harman, 2000). One series of workshops is the TREC-9 which was held at the National Institute of Standards and Technology in the year 2000. The previous eight TRECs were based on ad hoc search tasks unlike TREC-9 which focusses on seven tracks, that is cross-language retrieval, filtering, interactive retrieval, query analysis, question answering, spoken document retrieval and web retrieval. The current study employed the WT10G test collection, which is a 10 GB subset of TREC-9 focussing on the web track mimicking the retrieval environment of the World Wide Web. The topics covered in the web track were numbered from 415 to 500. Each topic in TREC consists of a title, a short description, and a narrative that spells out what would constitute a relevant article. Figure 3 illustrates a sample for topic 451 from the web track. In all, 20 topics were selected randomly for the experimental evaluation in this study.

TREC increases the time and the difficulty of building information retrieval systems, however the reliability of the experiment results are of no doubt as it provides a huge testing document collection, search topics and relevance judgments which truly simulate the real search environment (Kowalski and Maybury, 2002). Furthermore, TREC has been widely used to evaluate the performance of various retrieval systems (Xu *et al.*, 2010; Lagun and Agichtein, 2011).

The search engine

A prototype search engine was specifically built to facilitate users' task performance (i.e. query search), and to record all the key experimental measures (i.e. feedback). The following describes the CoRRe mechanism using sample screen grabs from the search engine.

Assume a query search is performed for the keyword "Australia." The system presents a list of initial results as illustrated in Figure 4. When the user places his/her mouse over the desired link, a snippet containing more information on the document appears. The options to provide explicit feedback are at the right-hand placement of the interface.

```
<num> Number:  451
<title> What is a Bengals cat?

<desc> Description:
Provide information on the Bengal cat breed.
<narr> Narrative:
Item should include any information on the Bengal cat breed, including description, origin,
characteristics, breeding program, names of breeders and catteries carrying bengals.
References which discuss bengal clubs only are not relevant.  Discussions of bengal tigers
are not relevant.
```

Figure 3.
An example topic for TREC-9

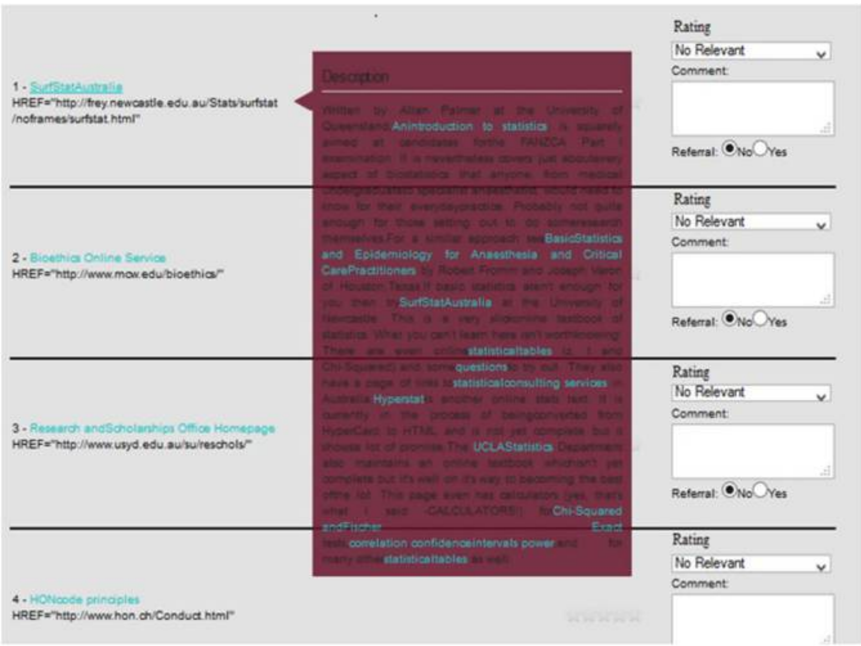

Figure 4.
A sample list of initial results for a query

Figure 5 shows the feedback provided by the user. Document no. 2 has been rated "Low relevant" and "Yes" for referral. On the other hand, Document no. 3 is rated as "Very relevant," commented as "Good" and has a "Yes" for referral. It is to note that for a new search as in this example, none of the documents have been previously rated, as indicated by the grey star(s).

Using the proposed re-ranking algorithm, the system re-ranks the original results based on the feedback provided by the user. Figure 6 shows the improved results for the same query, with Document no. 3 at the top most position whereas Document no. 1 has dropped to position number three. For clarity purpose, the original positions are indicated in blue. Moreover, documents containing feedback have their relevance shown in yellow star(s), therefore users can gauge the importance of the documents based on these star ratings. The top-5 documents will then be stored in the case base for future retrieval.

User evaluations

In all, 28 students with Computer Science backgrounds were recruited to evaluate CoRRe. Fifteen of the participants were males ($M_{age} = 22.4$; $SD_{age} = 0.89$) and the remaining 13 were females ($M_{age} = 21.7$; $SD_{age} = 1.02$). The experiment was conducted in a laboratory in the following manner:

- The research aims and the purpose of the experiment were briefly explained to the students.
- A list of 20 pre-fixed queries was provided to the students, based on the keywords. This is to ensure uniformity of the documents being searched. Some of

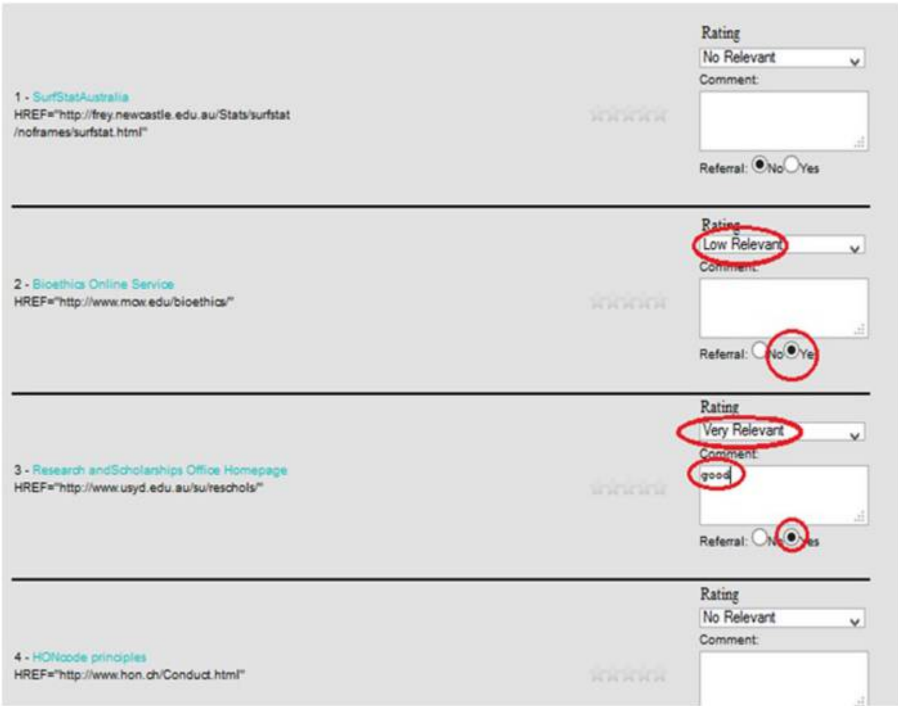

Figure 5.
A sample of explicit feedback

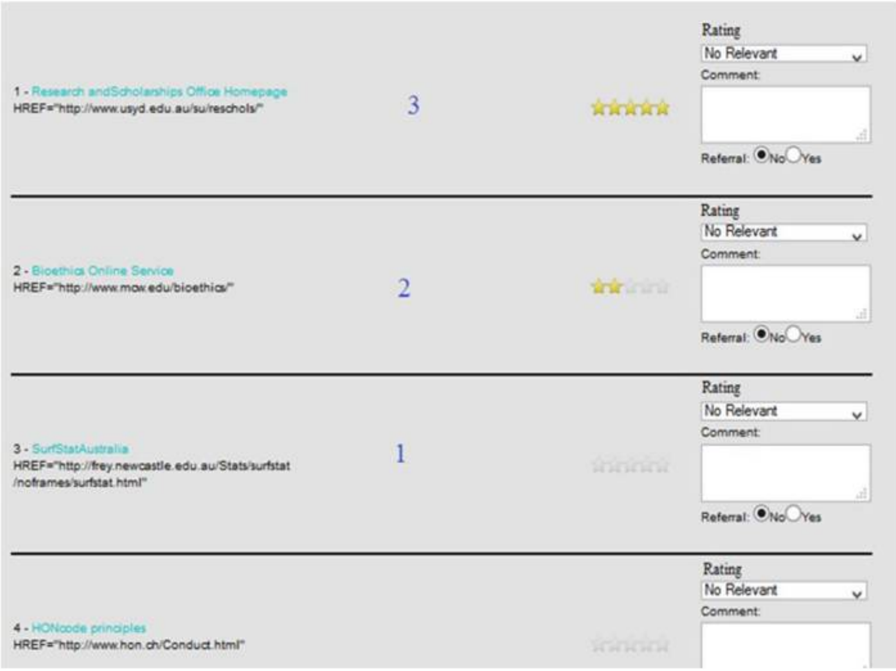

Figure 6. Improved results with star ratings

the keyword-based queries are "composition of zirconium," "world," "Parkinson's disease" and "Bengals," among others. The participants were encouraged to perform the search in no particular order based on the keywords, but they were required to perform all the searches at least once. Queries that have been searched at least once were then checked ("ticked") by the participants so that none is missed. In this manner, we ensured all the participants have performed all the queries given. Nevertheless, the participants were also encouraged to repeat some the queries, whenever appropriate.

A demo was provided on how to perform a query search, view the results and provide the explicit feedback. As the overall mechanism of the search engine is similar to other engines such as Google and Yahoo, the participants faced no problems in learning to use the system. Furthermore, the user interface was kept simple (see Figure 4), therefore the participants learned to use the system quickly. The experiments were conducted over a span of three weeks (i.e. once per week) with the participants spending approximately 30 minutes in each session. This was deemed important so that the participants do not get bored performing the searches, which may then result in them providing feedback inaccurately. Moreover, they were also left on their own without being monitored by the researcher. This was done to ensure the participants are comfortable, and to create a "normal" search environment. All the 28 participants were students who were registered for the Statistics course, therefore, the researcher carried out the experiments at the end of the weekly lecture sessions.

The system captures users' feedback automatically, which were then extracted to calculate its relevance. For instance, if User A provides rating and referral to Document no. 1, then this information (i.e. along with his/her rating and if the document has been referred or not) are captured, and used to re-rank the results. It is to note that the feedback details were captured separately, that is, rating for User A is stored separately from the referral. This process simplifies the calculations of MAP and NDCG, and comparisons between the models can be made easily. More details on these models are presented in the next section.

Evaluation metrics

Two evaluation metrics, namely, the MAP and NDCG were used to assess the efficacy of CoRRe. MAP has a good discrimination and stability compared to other measures (Salton and Buckley, 1988), and has been widely used in many retrieval studies (Balakrishnan and Zhang, 2014; Buscher *et al.*, 2012; Xu *et al.*, 2010; Agichtein *et al.*, 2006). A second evaluation was administered as a mean of comparison using NDCG, a more recent metric which handles multiple levels of relevance (Sakai, 2006; Bidoki *et al.*, 2010; Balakrishnan and Zhang, 2014; Xu *et al.*, 2010). NDCG basically measures the retrieval performance from the user's viewpoint and illustrates that the higher-ranking documents are more important to users compared to the lower-ranking documents.

Both the metrics were used to evaluate the retrieval precisions at top-5, 10 and 15 document levels. The document levels were standardized for both the metrics so that direct comparisons between the models can be made using MAP and NDCG. It is to note that when top-*k* levels were used to evaluate the performance of the models, only the top-*k* documents were taken into consideration, even if the results produced more relevant documents. In other words, if a query search produced tens or hundreds of results, only the top-5 documents were considered for the MAP and NDCG evaluations for a top-5 evaluation. Similarly, only the first ten and 15 documents were considered for top-10 and top-15 evaluations, respectively. This technique is usually implemented in information retrieval studies, as it has been shown that users generally only look at the top few results for each query search (Joachims *et al.*, 2005; Gao *et al.*, 2009).

Most of the studies in the information retrieval literature also compared proposed techniques with a baseline, which is usually a model without any feedback (White and Buscher, 2012; Ahn *et al.*, 2008; Balakrishnan and Zhang, 2014). As the greedy algorithm works without any user feedback, it is considered as the baseline model in the current study. The models evaluated and compared are as follows:

(1) baseline – model with no user feedback;

(2) rating – model incorporating users' ratings;

(3) comment – model incorporating users' comments;

(4) referral – model incorporating users' referrals; and

(5) CoRRe – the integrated feedback model.

In addition, independent sample *t*-tests were also administered to determine if the precision differences are significant between the various models. The difference(s) is considered to be significant at $p < 0.05$.

Results and discussion

Feedback analysis

A simple observation was made to assess the popularity of the feedback techniques among the participants in the study. The analysis shows that all the participants actively rated the documents, and approximately 65 percent of them actively used the referral feedback. The comments, however, received the lowest response rates, at approximately 57 percent. This is in line with issues related to explicit feedback whereby users are generally not keen in providing comments as it is more tedious, and consumes more time compared to clicking on the rating or the referral buttons (Ricci *et al.*, 2011; Raman *et al.*, 2012; Jung *et al.*, 2007; Hopfgartner and Jose, 2007; Xu *et al.*, 2010). But nevertheless, as will be demonstrated in the following sections, comments produce a better retrieval accuracy than referrals. This also shows that although referral is not a common feature, including on existing websites and recommender systems, the participants actively provided feedback via this mechanisms, probably due to it being a new idea (i.e. document retrievals) or simply because it is less tedious than commenting.

Baseline vs explicit feedback models

Table I illustrates the evaluation results for MAP for all the models used in the study.

Looking at Baseline and the rest of the models, it can be observed that the Baseline model had the lowest retrieval precisions compared to the feedback models, at all the document levels except in one scenario (i.e. Baseline vs Referral at top-5). However, the pattern changed when more documents were considered, that is, at top-10 and 15 with Referral outperforming Baseline.

In particular, huge differences can be noted between CoRRe and Baseline, ranging between 0.39 and 0.44. An independent sample *t*-test revealed the differences to be significant at all three document levels: top-5 ($t_{(18)} = -12.22, p < 0.001$), top-10 ($t_{(18)} = -8.035, p < 0.001$) and top-15 ($t_{(18)} = -7.05, p < 0.001$). Similar results were observed for NDCG, whereby CoRRe was found to outperform Baseline, as exemplified in Figure 7.

NDCG precisions were found to be higher than MAP, with CoRRe performing better than Baseline, regardless of the document levels. Independent sample *t*-test revealed these differences to be statistically significant as well – top-5 ($t_{(18)} = -3.049, p = 0.016$), top-10 ($t_{(18)} = -5.158, p = 0.001$) and top-15 ($t_{(18)} = -5.344, p = 0.01$). Therefore, based on MAP and NDCG results, CoRRe is considered a success when benchmarked against the Baseline model.

Model	Top-5	Top-10	Top-15
Baseline	0.237	0.241	0.243
Comment	0.273	0.315	0.386
Rating	0.424	0.426	0.431
Referral	0.212	0.256	0.272
CoRRe	*0.627*	*0.663*	*0.681*

Table I. Mean average precisions for the various models

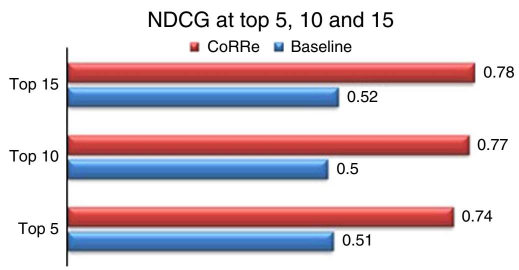

Figure 7. NDCG evaluations between CoRRe and Baseline

Comparison among the explicit feedback models
Looking across Table I, it can also be noted that CoRRe outperformed all the other models, at all the three levels. The improvements between CoRRe and the rest of the models are depicted in Figure 8.

From the figure, it is obvious that CoRRe showed great improvements over Referral, followed by Comment and Rating. In addition, the NDCG evaluations between CoRRe and the single feedback models also indicate CoRRe to outperform CoRRe as exemplified in Figure 9.

In fact, this result tallies with our second comparison, that is, between CoRRe. Results from Table I indicate Rating produced the best precision compared to Comment and Referral. Independent sample *t*-tests showed the differences to be significant between Rating and Comment at top-5 ($t_{(18)} = -8.12$, $p < 0.001$), top-10 ($t_{(18)} = -7.85$, $p = 0.031$) and top-15 ($t_{(18)} = -4.63$, $p = 0.043$). Similar results were obtained when statistical tests were carried out between Rating and Referral, with Rating significantly improving the retrieval accuracies at all three document levels: top-5 ($t_{(18)} = -9.11$, $p < 0.001$), top-10 ($t_{(18)} = -8.05$, $p = 0.002$) and top-15 ($t_{(18)} = -5.98$, $p = 0.0031$). The Comment model was also found to significantly improve the retrieval accuracy compared to Referral at all three document levels: top-5 ($t_{(18)} = -5.12$, $p = 0.012$), top-10 ($t_{(18)} = -5.92$, $p = 0.011$) and top-15 ($t_{(18)} = -3.98$, $p = 0.021$).

Discussion

The current study was carried out with the main aim of improving document retrieval relevancy by manipulating users' explicit feedback, namely, CoRRe. CoRRe was evaluated using two popular and established metrics, that is, MAP and NDCG. Findings from the study demonstrate improved retrieved accuracies through the use of relevancy feedback, particularly explicit feedback. These are now further elaborated.

First, our results indicate the Baseline model to have the lowest retrieval precisions compared to all the feedback models, except in one scenario (Baseline vs Referral at top-5).

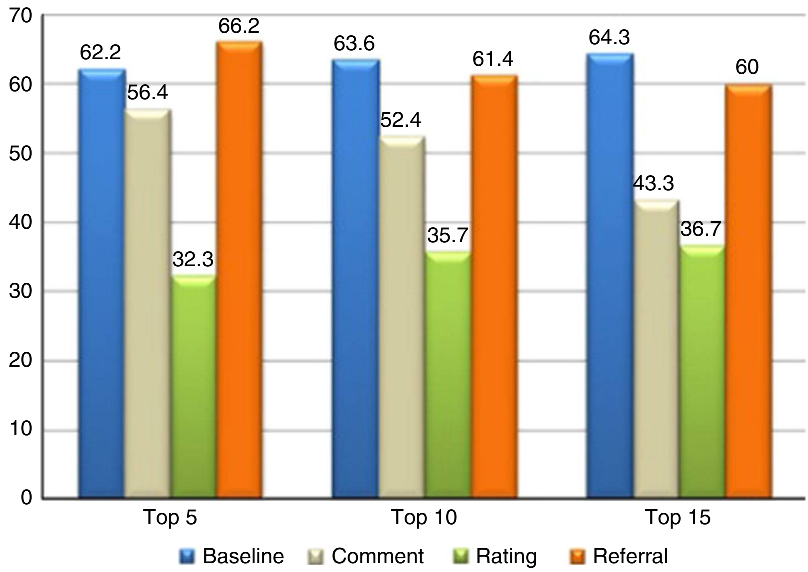

Figure 8.
Total MAP improvements (%) between CoRRe and the feedback models

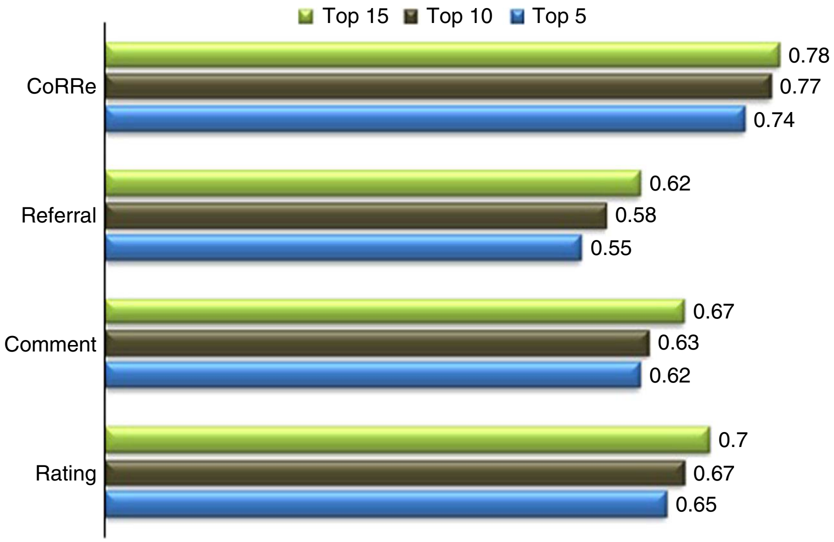

Figure 9.
NDCG between CoRRe and the single feedback models

This is probably due to the fact that only five documents were considered in the accuracy evaluations, therefore the baseline model performed slightly better than Referral. A lack of referrals from the participants for the top-5 documents may also be a cause. However, the performance of the Referral model improved over Baseline when more documents were considered (i.e. top-10 and 15). In general, findings show document retrieval performances to improve with the availability of users' feedback, and therefore concurring with previous studies that have examined both implicit feedback (White and Buscher, 2012; Buscher *et al.*, 2012; Balakrishnan and Zhang, 2014) and/or explicit feedback (Lagun and Agichtein, 2011; Lagun *et al.*, 2013; Tyler *et al.*, 2010; Takimoto, 2010). It is clear that users' feedback can be used successfully to improve retrieval relevancy, and thus enhance users' search experience as well. It is becoming a common practice for users to provide judgments or feedback online, for instance, on Amazon.com, eBay and Goodreads.com (books), to name a few. As explicit feedback are deemed to contain richer information (Buscher *et al.*, 2012; Jung *et al.*, 2007), then operators of recommender systems and search engines should look into tapping further into this feedback in order to infer users' interests, and provide more relevant results to the users, and thus improve users' search experience and satisfaction, as well.

Second, both evaluation metrics show that CoRRe outperformed all the single feedback models, and thus proving users' explicit thoughts can be integrated and used effectively to significantly improve the relevancy of the retrieved results. More specifically, this shows that CoRRe can be integrated successfully to improve retrieval relevancy. Generally, single feedback techniques are insufficient to predict users' interests, and are often unreliable (Bidoki *et al.*, 2010; Jung *et al.*, 2007; Claypool *et al.*, 2001), therefore it has become common for researchers to integrate numerous feedback techniques to improve retrieval results or to assess users' behavior and interests (Zhu *et al.*, 2012; Fox *et al.*, 2005; Liu *et al.*, 2011; Guo and Agichtein, 2012; Bidoki *et al.*, 2010). However, not all feedback approaches work in tandem, as conflicting results have been reported by many (Claypool *et al.*, 2001; Kelly and Teevan, 2003; Fox *et al.*, 2005). As our results indicate CoRRe to produce the best precisions, it has been demonstrated that users' comments, ratings and referrals can be effectively integrated to improve document relevance.

Third, a comparison across the three explicit feedback models revealed Rating to have produced the highest precision, followed by Comment and Referral. One of the main reasons for this is probably due to its popularity. The rating system is widely used by many commercial and non-commercial websites, therefore users tend to be more familiar and comfortable in providing the ratings (Buscher *et al.*, 2012). There exist various forms of ratings, ranging from five-point scales (e.g. Amazon.com, Goodreads.com) to unary scales (e.g. Twitter and Facebook). In fact, recent studies have started exploring the possibility of inferring users' interests based on the number of "Likes" received (Wang *et al.*, 2014), or more interestingly in determining users' personalities by exploring the "Likes" (Kosinski *et al.*, 2013; Bachrach *et al.*, 2012), suggesting the important roles of user ratings. The pattern of familiarity with ratings was also observed among our participants as all of them actively rated the documents based on their perception of relevancy. Additionally, unlike binary ratings (i.e. like or dislike) or unary ratings (i.e. like), CoRRe offers four varying levels for users to judge the relevancy of the documents, hence providing a better way for the users to rate their judgments. This also would have contributed to the higher precisions for Rating as users' feedback can be provided more accurately based on the options provided in the scale (i.e. ranging from no relevance to very relevant).

Text comments, which contain rich and more accurate representation of users' thoughts, emerged as the second best explicit feedback model. Comments are gaining popularity among the internet users, with many websites providing platforms for users to provide their thoughts, especially reviews on products and services (e.g. Amazon.com, eBay). As a matter of fact, business entities set up Facebook pages as a mean to promote their products and services (e.g. Emirates Airlines, McDonalds, etc.), and also to gauge users' perceptions. It is common to see Facebook users providing their opinions about a particular product or service on these pages. Other websites such as TripAdvisor not only asks for user ratings, but also provides a platform for users to comment about their experiences. Unlike Rating and Referral, Comment allows the users to explicitly express their opinions using textual comments (i.e. words) and therefore, they are able to portray their feelings (i.e. positive or negative) more accurately than the former techniques. It is also to note that the re-ranking algorithm used in this study analyzes the text comments provided by the users based on the nature of the comments, that is, positive or negative before re-ranking the search results. Furthermore, a higher weight is also assigned to Comment considering that it contains more meaningful representation of users' interests and also the rarity in users providing comments. It is believed that these two factors resulted in good retrieval precisions for the Comment model. In fact, analyzing comments based on sentiments (i.e. positive or negative) has been conducted successfully by mining users' opinions to improve users' decision-making process (Xu *et al.*, 2010; Liu, 2012; Shamim *et al.*, 2014).

The Referral model was introduced in this study as another form of feedback in which users recommend a document based on its relevancy. Referring a relevant document to the others is uncommon, except for sites such as Google+, which allows users to recommend interesting contents to users' contacts. In the current study, Referral works in a similar manner with Rating, in the sense that both require users to click on a button. However, unlike the varying levels of relevance, Referral was measured using a binary (i.e. Yes or No) scale, which translates to users referring a document if it is perceived to be highly relevant, otherwise the document is considered irrelevant. As stated previously, Referral was also found to be more popular among our participants compared to Comment, however the model produced lower precisions than Comment. One probable reason could be due to the weights assigned in the re-ranking formula. Although Rating and Referral were assigned a value of "1," the occurrences of ratings were much higher than the latter. On the other hand, Comment carried more weights than the other two models, therefore it produced better precisions than Referral. In other words, Referral performed poorly than Rating due to its lower frequencies, and poorly than Comment due to its lower relevance weights. Looking from another angle, it can also be argued that the concept of referring a relevant document is rather new to users, especially in information retrieval systems. Therefore, the feedback technique is not as popular as Rating among the participants. Even though the feedback gathered for Comment were the least, but this is probably due to "laziness in providing comments," and not because of unfamiliarity with the feedback technique. Nevertheless, Referral is still considered as a good feedback as it performed better than the Baseline model, and also worked well when integrated with Rating and Comment.

Overall the study demonstrated that users' CoRRe can be integrated to improve retrieval relevancies. Although it is common to see websites offering platforms for users to rate or comment, however not many include referral especially in information retrieval systems. Moreover, websites and recommender systems gather user comments, however these are not used to (re)-rank search results. Our findings indicate Comment to significantly improve retrieval results; hence e-commerce entities should capitalize on these findings and find ways to use users' explicit thoughts to improve overall users' search experiences. Finally, other than the three good feedback techniques, we also believe that the use of CBR may have simplified the whole process of retrieving the top most relevant documents to the users (i.e. top-5), hence improving the retrieval precisions and the overall efficiency of the retrieval system.

Conclusion, limitations and future work

The present study aimed to improve search results by manipulating users' explicit feedback. This was achieved by developing CoRRe, integrating three types of feedback, that is, users' comments, ratings and referrals. A re-ranking algorithm that ranks users' original results based on the five-star rating algorithm was proposed, and tested using the TREC document collection. Results indicate CoRRe to outperform all the models compared in the study, suggesting that although explicit feedback can be laborious and time consuming, they however can act as reliable indicators for relevancy. The combination of users' rich feedback, namely, comments, ratings and referrals provide significant advantages over using the feedback techniques alone as an indicator of relevance. In addition, rating seems to be the most popular feedback preferred by the users compared to referral and comments. The use of CBR also made it possible for the relevant results to be stored and retrieved efficiently, thus more information retrieval studies should look into using this technique to improve search results.

The study however, is not without its limitations. First, the weights for each of the explicit feedback techniques were assigned based on our understandings and assumptions. Future studies should further investigate this aspect to discover better ways in assigning the weights for each of the feedback, and then to evaluate their impacts on retrieval precisions. Second, although CBR is believed to have enhanced the retrieval process, the time taken to retrieve the documents based on a search query was not measured. This would have provided an indicator to the improvements gained by using CBR, hence future studies could look into measuring the retrieval time. By using two scenarios (i.e. with and without CBR), it would be easier to gauge the effect of CBR on the overall retrieval process.

Third, the study looked into integrating Comment, Rating and Referral to improve document retrieval results, and experiments were conducted to compare the integrated model with each of the explicit feedback techniques. The study however, did not examine the relationships between these feedback techniques, for instance, the correlations between users' ratings and referral. Although generally it can be safely assumed that the higher the rating, the higher the chances of a document to be recommended, it would still be interesting for future studies to look into these relationships.

Fourth, all the experiments conducted in this study focussed on the explicit feedback model, in line with our aim. However, the performance of CoRRe could be better established if it is also compared with other feedback models, such as implicit and pseudo-feedback models. There are many implicit feedback techniques such as, scrolling, dwell time and clicks, etc., therefore future studies could look into comparing CoRRe with these models (Claypool *et al.*, 2001; Fox *et al.*, 2005; White and Buscher, 2012), or with similar integrated models (Balakrishnan and Zhang, 2014; Lagun *et al.*, 2013).

Finally, the study implemented the greedy algorithm as the Baseline model. Although the algorithm is used in many information retrieval studies (Costa, 2010; Bhatt and Rusiya, 2013; Kumar and Sandeep,2013), it would also be interesting to use other improved baselines proposed by some researchers. For example, Lim *et al.* (2012) improved document retrievals by taking adjacent keywords into consideration (i.e. the order of the keywords). Future studies therefore could attempt to compare CoRRe or similar feedback models with such baselines.

References

Aamodt, A. and Plaza, E. (1996), "Case-based reasoning: foundational issues, methodological variations, and systems approaches", *Artificial Intelligence Communications*, Vol. 7 No. 1, pp. 39-59.

Agichtein, E., Brill, E. and Dumais, S. (2006), "Improving web search ranking by incorporating user behaviour information", *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, August 6-11*, pp. 19-26.

Ahn, J., Brusilovsky, P., He, D., Grady, J. and Li, Q. (2008), "Personalized web exploration with task models", *Proceeding of the 17th International Conference on World Wide Web 2008, April 21-25*, pp. 1-10. doi: 10.1145/1367497.1367499.

Amatriain, X., Pujol, J., Tintarev, N. and Oliver, N. (2009), "Rate it again: increasing recommendation accuracy by user rerating", *Proceedings of the 3rd ACM Conference on Recommender Systems*, pp. 173-180.

Anderson, M. (2014), "88% of consumers trust online reviews as much as personal recommendations", available at: http://searchengineland.com/88-consumers-trust-online-reviews-much-personal-recommendations-195803 (accessed September 2015).

Bachrach, Y., Kosisnky, M., Graepel, T., Kohli, P. and Stillwell, D. (2012), "Personality and patterns of Facebook usage", Web Science'12, Evanston, IL.

Balakrishnan, V. and Zhang, X.Y. (2014), "Implicit user behaviours to improve post-retrieval document relevancy", *Computers in Human Behavior*, Vol. 33, pp. 104-112.

Baeza-Yates, R. and Ribeiro-Neto, B. (1999), *Modern Information Retrieval*, ACM Press, London.

Beebe, N.L., Clark, J.G., Dietrich, G.B., Ko, M.S. and Ko, D. (2011), "Post-retrieval search hit clustering to improve information retrieval effectiveness: two digital forensics case studies", *Decision Support Systems*, Vol. 4 No. 4, pp. 732-744.

Belkin, N.J., Cool, C., Stein, A. and Thiel, U. (1995), "Cases, scripts, and information-seeking strategies: on the design of interactive information retrieval systems", *Expert Systems with Applications*, Vol. 9, pp. 379-395.

Bhatt, P. and Rusiya, P. (2013), "GRF: greedy based relevance feedback algorithm for retrieval of multimedia object", *International Journal of Computer Science and Information Technologies*, Vol. 4 No. 4, pp. 649-650.

Bidoki, Z.A.M., Ghodsnia, P., Yazdani, N. and Oroumchian, F. (2010), "A3CRank: an adaptive ranking method based on connectivity, content and click-through data", *Information Processing and Management*, Vol. 46 No. 2, pp. 159-169.

Bixby, D.L. (2013), "Managing inadequate responses to frontline treatment of chronic myeloid leukemia: a case-based review", *Cancer Treatment Reviews*, Vol. 39 No. 3, pp. 241-251.

Bullock, B.N., Robert, J. and Andreas, H. (2011), "Tagging data as implicit feedback for learning-to-rank", *Proceedings of the ACM Web Science Conference*, *New York, NY, June 14-17*, pp. 1-4.

Buscher, G., White, R.W., Dumais, S. and Huang, J. (2012), "Large-scale analysis of individual and task differences in search result page examination strategies", *Proceedings of the fifth ACM International Conference on Web Search and Data Mining, Feburary 8-12*, pp. 373-383.

Chen, M. (2008), "A greedy algorithm with forward looking strategy", in Bednorz, W. (Ed.), *Greedy Algorithms*, InTech Publishers, Vienna, pp. 1-16.

Choo, E., Yu, T., Chi, M. and Sun, Y.L. (2014), "Revealing and incorporating implicit communities to improve recommender systems", *Proceedings of the fifteenth ACM Conference on Economics and Computation, June 8-12*, pp. 489-506.

Claypool, M., Le, P., Wased, M. and Brown, D. (2001), "Inferring user interest", *IEEE Internet Computing*, Vol. 5 No. 6, pp. 32-39.

Costa, A. (2010), "Optimization algorithm for improving the efficacy of an information retrieval model", *Proceedings of Toulouse Global Optimization 2010 (TOGO), August 31-September 3*, pp. 1-4.

Daniels, J.J. and Rissland, E.L. (1995), "A case-based approach to intelligent information retrieval", *Proceeding SIGIR' 95 Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, July 9-13*, pp. 238-245.

Dasgupta, S., Papadimitriou, C.H. and Vazirani, U.V. (2006), "Algorithms", available at: http://beust.com/algorithms.pdf (accessed November 2014).

Fox, S., Karnawat, K., Mydland, M., Dumais, S. and White, T. (2005), "Evaluating implicit measures to improve web search", *ACM Transactions on Information Systems*, Vol. 23 No. 2, pp. 147-168.

Gao, J., Yuan, W., Li, X., Deng, K. and Nie, J.-Y. (2009), "Smoothing clickthrough data for web search ranking", *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval – SIGIR, July 19-23*, pp. 355-362.

Garcia Esparza, S., O'Mahony, M.P. and Smyth, B. (2012), "Mining the real-time web: a novel approach to product recommendation", *Journal of Knowledge-Based Systems: Special Issue on Innovative Techniques and Applications of Artificial Intelligence*, Vol. 29, pp. 3-11.

Guo, Q. and Agichtein, E. (2012), "Beyond dwell time: estimating document relevance from cursor movements and other post-click searcher behavior", *Proceedings of the 21st International Conference on World Wide Web, April 16-20*, pp. 569-578.

He, W., Erdelez, S., Wang., F. and Shyu, C. (2008), "The effects of conceptual description and search practice on users' mental models and information seeking in a case-based reasoning retrieval system", *Information Processing and Management: An International Journal*, Vol. 44 No. 1, pp. 294-309.

Henriet, J., Lenia, P.E., Laurenta, R. and Salomonb, M. (2014), "Case-based reasoning adaptation of numerical representations of human organs by interpolation", *Expert Systems with Applications*, Vol. 41 No. 2, pp. 260-266.

Hopfgartner, F. and Jose, J. (2007), "Evaluating the implicit feedback models for adaptive video retrieval", *Proceedings of the 9th ACM SIGMM International Workshop on Multimedia Information, September 28-29*. doi: 10.1145/1290082.1290127.

Hu, M. and Liu, B. (2004), "Mining and summarizing customer reviews", *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004), Seattle, Washington, DC, August 22-25*.

Jawaheer, G., Szomszor, M. and Kostkova, P. (2010), "Comparison of implicit and explicit feedback from an online music recommendation service", *Proceedings of the 1st International Workshop on Information Heterogeneity and Fusion in Recommender Systems, September 26*, pp. 47-51.

Jawaheer, G., Weller, P. and Kostkova, P. (2014), "Modeling user preferences in recommender systems: a classification framework for explicit and implicit user feedback", *ACM Transactions on Interactive Intelligent Systems*, Vol. 4 No. 2, pp. 1-26.

Joachims, T., Granka, L., Pan, B., Hembrooke, H. and Gay, G. (2005), "Accurately interpreting clickthrough data as implicit feedback", *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval – SI*, pp. 154-161.

Jung, S., Herlocker, J.L. and Webster, J. (2007), "Click data as implicit relevance feedback in web search", *Information Processing and Management*, Vol. 43 No. 3, pp. 791-807.

Kelly, D. and Teevan, J. (2003), "Implicit feedback for inferring user preference: a bibliography", *ACM SIGIR Forum*, Vol. 37 No. 2, pp. 18-28.

Kim, H. (2014), "Retrieval effectiveness of controlled and uncontrolled index terms in INSPEC database", *Malaysian Journal of Library & Information Science*, Vol. 19 No. 2, pp. 103-117.

Kimura, A., Kashino, K., Kurozumi, T. and Murase, H. (2008), "A quick search method for audio signals based on a piecewise linear representation of feature trajectories", *IEEE Transactions on Audio Speech Language Process*, Vol. 16 No. 2, pp. 396-407.

Kolodner, J.L. (1999), "An introduction to case-based reasoning", *Artificial Intelligence Review*, Vol. 6 No. 1, pp. 3-34.

Kosinski, M., Stillwell, D.J. and Graepel, T. (2013), "Private traits and attributes are predictable from digital records of human behavior", in Wachter, K. (Ed.), *Proceedings of the National Academy of Sciences of USA*, Vol. 110, US National Academy of Science, Berkeley, CA, pp. 5802-5805. doi:10.1073/pnas.1218772110.

Kowalski, G.J. and Maybury, M.T. (2002), *Information Storage and Retrieval Systems: Theory and Implementation*, Kluwer Academic, Boston, MA.

Kumar, M.A. and Sandeep, Y. (2013), "Clustering analysis based on greedy heuristic algorithm", *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol. 3 No. 11, pp. 1332-1334.

Lagun, D. and Agichtein, E. (2011), "ViewSer: enabling large-scale remote user studies of web search examination and interaction categories and subject descriptors", *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, July 24-28*, pp. 365-374.

Lagun, D., Sud, A., White, R.W., Bailey, P. and Buscher, G. (2013), "Explicit feedback in local search tasks", *SIGIR' 13 Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, July 28-August 1*, pp. 1065-1068.

Lerman, K. (2006), "Social networks and social information filtering on digg", available at: http://arxiv.org/pdf/cs/0612046.pdf (accessed September 2015).

Lim, B.H., Balakrishnan, V. and Ram, G.R. (2012), "Improving the relevance of document search using the multi-term adjacency keyword-order model", *Malaysian Journal of Computer Science*, Vol. 25 No. 1, pp. 1-10.

Liu, B. (2012), *Sentiment Analysis and Opinion Mining*, Morgan and Claypool Publishers, CA.

Liu, Y., Miao, J., Zhang, M., Ma, S. and Ru, L. (2011), "How do users describe their information need: query recommendation based on snippet click model", *Expert Systems with Applications*, Vol. 38 No. 11, pp. 13847-13856.

Lu, Y., Zhai, C.X. and Sundaresan, N. (2009), "Rated aspect summarization of short comments", *Proceedings of the 18th International Conference on World Wide Web (WWW'09), ACM Press, New York, NY*, pp. 131-140.

Lupiani, E., Juarez, J.M. and Palma, J. (2014), "Evaluating case-base maintenance algorithms", *Knowledge-Based Systems*, Vol. 67, pp. 180-194.

Manning, C.D., Raghavan, R. and Schutze, H. (2009), *An Introduction to Information Retrieval*, Cambridge University Press, New York, NY.

Masuda, T., Yamamoto, D., Ohira, S. and Nagao, K. (2008), *Video Scene Retrieval Using Online Video Annotation", Lecture Notes on Artificial Intelligence*, Springer-Verlag, Berlin, Heidelberg, pp. 54-62.

Mudambi, S.M. and Schuff, D. (2010), "What makes a helpful online review? A study of customer reviews on amazon.com", *MIS Quarterly*, Vol. 34 No. 1, pp. 185-200.

Ngo-Ye, T.L. and Sinha, A.P. (2012), "Analyzing online review helpfulness using a regressional relief enhanced text mining method", *ACM Transactions on Management Information Systems*, Vol. 3 No. 2, pp. 1-20.

Núñez-Valdéz, E.R., Cueva Lovelle, J.M., Sanjuán Martínez, O., García-Díaz, V., Ordoñez de Pablos, P. and Montenegro Marín, C.E. (2012), "Implicit feedback techniques on recommender systems applied to electronic books", *Computers in Human Behavior*, Vol. 28 No. 4, pp. 1186-1193.

Oard, D.W. and Kim, J. (1998), "Implicit feedback for recommender systems", *Proceedings of the AAAI Workshop on Recommender Systems, July 26-27*, pp. 81-83.

Raman, K., Shivaswamy, P. and Joachims, T. (2012), "Online learning to diversify from implicit feedback", *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Beijing, August 12-16*. doi:10.1145/2339530.2339642.

Ricci, F., Rokach, L., Shapira, B. and Kantor, P.B. (2011), *Recommender Systems*, Springer, New York, NY.

Sakai, T. (2006), "On the reliability of information retrieval metrics based on graded relevance", *Information Processing and Management*, Vol. 43 No. 2, pp. 531-548.

Salton, G. and Buckley, S. (1988), "Term weighting approaches in automatic text retrieval", *Information Processing and Management*, Vol. 24 No. 5, pp. 513-523.

Shamim, A., Balakrishnan, V., Tahir, M. and Shiraz, M. (2014), "Critical product features' identification using an opinion analyzer", *The Scientific World Journal*, Vol. 2014, Article ID 340583, Vol. 2014, available at: http://dx.doi.org/10.1155/2014/340583

Siersdorfer, S., Chelaru, S., Nejdl, W. and Pedro, J.S. (2010), "How useful are your comments?: analyzing and predicting YouTube comments and comment ratings", *Proceedings of the 19th International Conference on World Wide Web (WWW'10), ACM Press, New York, NY*, pp. 891-900.

Sinha, R. and Swearingen, K. (2001), "Comparing recommendations made by online systems and friends", *Proceedings of the DELOS-NSF Workshop on Personalization and Recommender Systems in Digital Libraries, June 18-20*, available at: www.ercim.eu/publication/ws-proceedings/DelNoe02/RashmiSinha.pdf (accessed September 2015).

Takimoto, M. (2006), "The effects of explicit feedback on the development of pragmatic proficiency", *Language Teaching Research*, Vol. 10 No. 4, pp. 393-417.

Tyler, S.K., Wang, J. and Zhang, Y. (2010), "Utilizing re-finding for personalized information retrieval", *Proceedings of the 19th ACM International Conference on Information and Knowledge Management, October 26-30*, pp. 1469-1472.

Voorhees, E. and Harman, D. (2000), "Overview of the ninth text retrieval conference (TREC-9), National Institute of Standards and Technology", report, Gaithersburg, available at: http://trec.nist.gov/pubs/trec9/papers/overview_9.pdf (accessed February 2014).

Wakamiya, S., Kitayama, D. and Sumiya, K. (2011), "Scene extraction system for video clips using attached comment interval and pointing region", *Multimedia Tools Application*, Vol. 54 No. 1, pp. 7-25.

Wang, B., Ester, M., Bu, J. and Cai, D. (2014), "Who also likes it? Generating the most persuasive social explanations in recommender systems", *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27-31*, pp. 173-179.

Wei, P.S. and Lu, H.P. (2013), "An examination of the celebrity endorsements and online customer reviews influence female consumers' shopping behavior", *Computers in Human Behavior*, Vol. 29 No. 1, pp. 193-201.

White, R.W. and Buscher, G. (2012), "Text selections as implicit relevance feedback", *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, August 12-16*, pp. 1151-1152.

Xu, J., Chen, C., Xu, G., Li, H. and Abib, E. (2010), "Improving quality of training data for learning to rank using click-through data", *Proceedings of the Third ACM International Conference on Web Search and Data Mining – WSDM, February 4-6*, pp. 171-180.

Yan, A., Qian, L. and Zhang, C. (2014), "Memory and forgetting: an improved dynamic maintenance", *Information Sciences*, Vol. 287, pp. 50-60.

Zhu, Y., He, L. and Wang, X. (2012), "User interest modeling and self-adaptive update using relevance feedback technology", *Procedia Engineering*, Vol. 29, pp. 721-725.

Further reading

Duda, R.O. and Hart, P.E. (1973), *Pattern Classification and Scene Analysis*, Wiley, New York, NY.

Corresponding author
Dr Vimala Balakrishnan can be contacted at: vim.balakrishnan@gmail.com