Retrieval through explanation:
an abductive inference approach to relevance feedback

I. Ruthven, University of Glasgow
M. Lalmas, Queen Mary and Westfield College, University of London
C. J. van Rijsbergen, University of Glasgow
Email: igr, keith@dcs.gla.ac.uk, mounia@dcs.qmw.ac.uk

Abstract

Relevance feedback techniques are designed to automatically improve a system's representation of a query by using documents the user has marked as relevant. However, traditional relevance feedback models suffer from a number of limitations that restrict their potential in supporting information seeking. One of the major limitations of relevance feedback is that it does not incorporate behavioural aspects of information seeking - how and why users assess relevance. We propose that relevance feedback should be viewed as a process of explanation and demonstrate how this limitation of relevance feedback techniques can be overcome by a theory of relevance feedback based on abductive inference.

1. Introduction

Information retrieval (IR) systems are designed to retrieve documents that are relevant to a user's information need, which is usually expressed as a query. However, selecting good query terms to represent an information need is difficult. The complexity of verbalising an information need can increase when the need is vague [SW99], when the document collection is unfamiliar, [Roc71], or when the searcher is inexperienced with information retrieval (IR) systems, [CPBKN96]. It is much easier, however, for a user to assess which documents contain relevant information.

Relevance feedback (RF) techniques make use of this fact to automatically modify a query representation based on the documents a user considers to be relevant. RF is usually presented as a cycle of activity: an IR system presents a user with a set of retrieved documents, the user indicates those that are relevant and the system uses this information to produce a new query, which is then used to retrieve more documents. This cycle is known as an *iteration* of RF.

RF has proved to be relatively successful at increasing the effectiveness of retrieval systems in certain types of search, [Har92]. However, the traditional approaches to RF do not consider the *behavioural* aspects of information seeking and making relevance assessments. The standard RF algorithms consider only *which* documents a user has marked as relevant; they do not consider *why* a user has assessed relevance.

In addition, RF has to consider *how* a user assesses relevance. Current RF algorithms lack the qualitative reasoning ability necessary to tackle aspects of relevance assessments such as developing information needs [Cam95], consistency of relevance assessments [BI97] and a user's strategies for assessing relevance [FM95].

We view RF as a process of *explanation.* An RF theory should provide an explanation of why a document is relevant to an information need. These explanations can be based on how information is used within documents. We propose *abductive logic*, [JJ94] as a suitable framework for an explanation-based account of RF. Abductive inference provides hypotheses that explain a given set of data. For example, given that a patient has red spots, an abductive inference engine could provide hypotheses that the patient is suffering from chickenpox, measles or a number of other conditions. Each of these hypotheses are possible *explanations* of the patient's condition; some explanations being more likely than others depending on what information is available to generate the hypotheses.

In RF this framework produces a set of possible explanations for why a user marked a number of documents relevant at the current search iteration. From the set of possible explanations, one explanation, known as the *best possible explanation*, is selected to reformulate the query. Abductive logic allows us to incorporate information on how a user is searching for information, and why documents are relevant based on how information is used in the documents. In order to be able to utilise how information is used in documents we need to extend the representation, or *indexing*, language used by IR systems. We discuss this in section 2.

In the remainder of the paper we outline a model for RF based on abductive inference. We show in section 3 how this technique can be used to model RF and demonstrate how some behavioural aspects of searching can be described in an abductive framework. We conclude in section 4.

2. Expressiveness of indexing language

In [RL99] we described how the simple term-based approach to document indexing can be extended to include information on how individual terms are used within documents. We defined a number of weighting schemes - *term characteristics* - to describe aspects of the use of a term in a document, e.g. *theme* - a characteristic which describes the degree to which the term is likely to be the main topic of a document, *context* - a characteristic which describes the dependence of one query term on another. Each weighting scheme gives a weight to a term, reflecting some aspect of how a term is used within a document or collection. The measures we used in [RL99] were very simple estimations but can be replaced by more sophisticated algorithms to cope with, for example, very long documents [HP93], or documents for which we have more detailed identification of important concepts [PJ93].

The traditional approach to RF produces a new query, **(Q1),** consisting of a set of terms and frequency-based weights. In our approach, the new query, **(Q2),** includes information on each of the terms role in the relevant documents, for instance:

index(millennium)	0.5	*index(millennium)*	0.5,	*theme(millennium)*	0.3
index(computer)	0.4	*index(computer)*	0.4,	*theme(computer)*	0.7
index(bug)	0.1	*index(bug)*	0.1,	*context(millennium, bug)*	0.4
(Q1)		**(Q2)**			

A query then consists of a set of term characteristics rather than a set of terms. Query **(Q1)** would prioritise retrieval of any document containing the three terms, *millennium, computer*, and *bug*, whereas query **(Q2)** would prioritise retrieval of documents in which the main topics were *millennium* and *computer*, and the terms *millennium* and *bug* appeared in close proximity to each other.

In [RL99] we demonstrated that RF is not only a case of selecting *which* terms to use in a query but which *characteristics* of a term. We showed, experimentally, that it was possible to select, for a given set of relevance assessments, an optimal set of term characteristics to use in a query. This means different sets of term characteristics are better at retrieving relevant documents for different queries. Furthermore, we showed, through a series of experiments on large document collections, that this technique not only performed well, but outperformed classic RF algorithms such as the $F_{4.5}$ measure, [RSJ76].

Term characteristics provide the basis for an indexing language for our explanation-based account of RF. Each term characteristic indicates how a term is used within a document and can be used to differentiate why a user marked a document as relevant.

3. Incorporation of behavioural aspects of searching

Critics of RF techniques, e.g. [Cam95, SW99, DBM97], argue that the traditional RF techniques make too many simplifying assumptions about users' role in the RF process. For example, RF techniques traditionally assume that information needs are fixed, that all relevant documents are equally relevant and that only one information need is under investigation in a search. Studies of how users make relevance assessments, e.g.[FM95], show that this is not the case: documents may only be partially relevant, different documents may satisfy different aspects of an information need and those documents that users finds relevant may change during a search. Why a user assesses material as relevant should then be central to RF.

We believe that understanding relevance assessments should be a process of *explanation:* explaining why a user made a relevance assessment on a document. In this section we briefly outline abductive inference and discuss its suitability as a theory for RF. In section 3.2 we discuss factors that influence the abductive inference. We then show how this model can incorporate aspects of user behaviour such as dynamic information needs (section 3.3) and user uncertainty (section 3.4).

3.1. Abductive inference and RF

Abductive inference provides hypotheses that explain a given set of data. In RF, the data we are seeking to explain are the relevance assessments given by the user on the documents retrieved by the IR system, and an explanation consists of a set of term characteristics.

In the example below we have a query '*endowment mortgage*' and three relevant documents, **(D1)**, **(D2)** and **(D3)**. The documents are represented by the term characteristics of the query terms they contain.

theme(mortgage)	*theme(mortgage)* and *context(mortgage, endowment) and*	*theme(mortgage)* and *theme(endowment)* and *context(mortgage, endowment)*
(D1)	**(D2)**	**(D3)**

Possible explanations for marking these documents relevant include:

•*theme(mortgage)* - "relevant documents contain the term *mortgage* as a main topic",

•*context(mortgage, endowment)* - "relevant documents mention *mortgage* and *endowment* in close proximity",

•*theme(mortgage)* and *context(mortgage, endowment)* - "relevant documents have *mortgage* as a main topic but will also mention *endowment* in close proximity to *mortgage*".

Potentially any set of term characteristics can constitute an explanation. However, not all explanations will be equally good at explaining the relevance assessments. In addition to providing a set of possible explanations, abductive logic allows for the identification of the *best possible explanation.* The task is then to infer which is the best possible explanation for the set of relevant documents. The best explanation can then be used to provide a modified query to retrieve more documents. In the next section we discuss some of the factors of the IR system that influences the choice of best explanation.

3.2. System factors

There are be a number of factors to consider in deciding which is the best explanation for a set of relevance assessments. These include coverage of documents, uncertainty of events and the strength of a term characteristic.

Coverage of documents: In the previous example, *theme(endowment)* may be a poorer explanation than *theme(mortgage)* because *theme(endowment)* only explains the relevance of one of the three documents, whereas *theme(mortgage)* is an explanation of the relevance of all three documents.

Uncertainty of events: The process of generating abductive explanations and selecting the best possible explanation must support *uncertainty* in information seeking, [Kuh93]. The evidence given by terms and term characteristics, as to the relevance of documents, is by nature uncertain. The choice of best possible explanation is affected, amongst other things, by the likelihood of the characteristics that form the explanation. For example, a term characteristic that appears more frequently in relevant than non-relevant documents, those documents not explicitly marked relevant, is more likely to provide a good explanation.

Strength of term characteristic: In addition to how often a term appears in relevant and irrelevant documents, a term characteristic that is more strongly displayed in relevant documents - has a higher weight - is more likely to provide a good explanation.

The weights assigned to the characteristics of terms in documents and the frequency of characteristics of terms in relevant and irrelevant documents both help in modelling aspects of retrieval that affect the system. A good explanation is then one that explains most of the documents using characteristics of terms. Characteristics of terms that are useful for explanation display two features: they are more strongly displayed in the relevant than irrelevant documents and they appear more often in the relevant than irrelevant documents.

3.3. Dynamic information needs

Information needs can change over time: the user may be forced to change what information he is looking for because the collection contains little relevant information, the user may have started his search with a poorly-defined information need or the user may learn more about the topic on which he is searching during the search.

The result is that what constitutes a good query at each iteration of relevance feedback may vary. Traditional RF algorithms condense all relevant information into one set - they ignore the temporal nature of information seeking. We can incorporate this aspect of searching into an abductive framework in two ways:

i. by considering when a document was marked relevant, [Cam95]. A term or characteristic of a term which is good at describing relevance at the start of a search may not be equally good at describing relevance at the end of a search. We need, then, to infer how the search is changing in *topic* over time. In the following example, three documents were marked relevant (**D1** - **D3**), one at each iteration of RF (**I1** - **I3**):

theme(mortgage)	*theme(mortgage)* and	*theme(mortgage)* and
theme(pensions)	*theme(endowment) and*	*theme(tax)* and
		context(mortgage, endowment)
(D1)	**(D2)**	**(D3)**
(I1)	**(I2)**	**(I3)**

At **I3** *theme(mortgage)* can be regarded as a relatively fixed part of the information need as it appears in the relevant documents at each iteration. Other characteristics of terms, such as *theme(pensions)* and *theme(endowment),* only appear in one iteration and so should not be regarded as being as good as *theme(mortgage)* at indicating relevance at the current search stage. Thus we infer, in the absence of any other information, that *theme(mortgage)* is a better term characteristic to appear in an good explanation.

At **I3**, *theme(pension)* and *theme(tax)*, both appear in one relevant document but *theme(tax)* has appeared in the most recent relevant document (**D3**) whereas *theme(pension)* has appeared much earlier in the search. In this case we prefer the selection of an explanation that contains *theme(tax)* as it is more indicative of the information the user is currently assessing as relevant.

ii. by considering the consistency of relevance assessments at each iteration. Documents at each stage may be more or less consistent, or similar, in the term characteristics they display. High consistency of term characteristics can reflect a very focussed search, whereas low consistency can reflect a vague stage in a search or a search in which different documents reflect different aspects of the underlying information need [SW99]. Examining the consistency of documents at each iteration can help infer how searching is changing in focus - whether the search is becoming more refined or becoming broader in focus. This information can be used to choose the best explanation, e.g. a search that is becoming more focussed may require an explanation that is more specific.

The quality of an explanation, then, is not just how good it is at explaining the relevant assessments made at the current iteration but also how well it explains the assessments the context of a whole search. An inference model for generating explanations incorporates how a search is changing over time.

3.4 Partial relevance and user uncertainty

In section 3.2. we discussed how abductive explanations can support modelling of system factors. Uncertainty can also arise at the user level [Kuh93]. One way this can be demonstrated is in the use of partial relevance assessments, in which the user does not only mark documents as relevant but indicates how relevant the documents are.

At the simplest level we can incorporate partial relevance by modifying the weights attached to term characteristics (section 3.2). This views the relevance of a document as a variable to be handled by the system. However we can use the partial relevance scores, combined with the information on how the search is changing (section 3.3), to *direct* the selection of best possible explanation, e.g. in favour of an explanation that explains either most of the relevance assessments (if the search is very broad) or only the most relevant documents (if the search is very focussed).

4. Conclusion

In this paper we outline how a model of RF based on abductive inference can overcome some of the limitations of traditional RF algorithms in incorporating behavioural aspects of information seeking. The basic inference is that of selecting a good explanation - set of term characteristics - to modify a query but there is also inference related to how the searching is changing both in topic and focus.

We also stressed that this approach is one of selection - actively selecting a response to the user's actions based on the relevance assessments she makes. Traditional approaches to RF cannot easily incorporate this kind of reasoning because they rely on passive techniques to reformulate queries. The move towards this type of reasoning approach is necessary if RF is to take account of the phenomena described in studies of user searching behaviour.

References

We are grateful to Mark Dunlop, Joemon Jose and the anonymous referees for their help in improving this paper and their insightful comments on the work.

References

[BI97] P. Borlund and P. Ingwersen. The development of a method for the evaluation of interactive information retrieval systems. *Journal of Documentation*. 53 (5). pp 225 -250. 1997.

[Cam95] I. Campbell. Supporting information needs by ostensive definition in an adaptive information space. *MIRO '95*. electronic Workshops in Computing. Springer. I. Ruthven (ed). 1995.

[CPBKN96] C. Cool, S. Park, N. J. Belkin, J. Koenemann, K. B. Ng. Information seeking behavior in new searching environment. *CoLIS 2.* Copenhagen. pp 403-416. 1996.

[DBM97] N. Denos, C. Berrut and M. Mechkour. An image system based on the visualization of system relevance via documents. In: *Database and Expert Systems Applications (DEXA '97).* pp 214-224. Springer.1997.

[FM95] V. Florance and G. Marchionini. Information processing in the context of medical care. *Proceedings of ACM SIGIR Conference on Research and Development in Information Retrieval.* ACM Press.. Seattle. pp158 -163. 1995.

[JJ94] J. R. Josephson and S. G. Josephson (ed). *Abductive Inference: Computation, Philosophy, Technology.* Cambridge University Press. 1994.

[Har92] D. Harman. Relevance feedback and other query modification techniques. In: *Information retrieval: data structures & algorithms* . W. B. Frakes and R. Baeza-Yates, ed. pp 241-263. 1992

[HP93] M.A. Hearst and C. Plaunt. Subtopic structuring for full-length document access. *Proceedings of ACM SIGIR Conference on Research and Development in Information Retrieval.* Pittsburgh, USA. ACM Press. pp 59-68. 1993.

[Kuh93] C.C. Kuhlthau. Principle for uncertainty for information seeking. *Journal of Documentation.* 49(4). pp 339-355, 1993.

[PJ93] C. D. Paice and P. A. Jones. The identification of important concepts in highly structured technical papers. *Proceedings of ACM SIGIR Conference on Research and Development in Information Retrieval*. Pittsburgh, USA. ACM Press. pp 69-79. 1993.

[RSJ76] S. E. Robertson and K. Sparck Jones. Relevance weighting of search terms. *Journal of the American Society of Information Science.* 27. pp129-146.1976.

[Roc71] J. J. Rocchio. Relevance feedback in information retrieval. In: *The SMART retrieval system: experiments in automatic document processing.* G. Salton, ed. pp 313-323. Prentice-Hall.

[RL99] I. Ruthven and M. Lalmas. Selective relevance feedback using term characteristics. *CoLIS 3.* Dubrovnik. 1999.

[SW99] A. Spink, and T. D. Wilson. Toward a theoretical framework for information retrieval (IR) evaluation in an information seeking context. In *Mira '99.* electronic Workshops in Computing. S. Draper, M. Dunlop, I. Ruthven and C. J. van Rijsbergen (ed). *(to appear).*