Context Generation in Information Retrieval

Ian Ruthven and C J van Rijsbergen
Department of Computing Science
University of Glasgow, Glasgow, G12 8QQ
Scotland
email: igr,keith@dcs.gla.ac.uk

Abstract

The interaction between a user and an information retrieval system can be viewed as a dialogue in which both participants are trying to interpret the others' actions in the light of previous experience. The system then must try to generate a context in which to interpret the user's response to the presented material. This notion of context operates on a principle of relevance. Information that the system believes is relevant to the user, or that the user has indicated as relevant will form the basis of the system's notion of the context. This paper presents a way of representing a context that can use both the systems knowledge about itself and the user's response to generate a view of the retrieval session.

1 INTRODUCTION

The notion of relevance is central to Information Retrieval (IR). The system should return documents it assesses as being likely to be relevant to a user's request. This request traditionally takes the form of a query. A more interesting concept is that of pertinence, [Kem74], a definition based on the utility of the information contained within the documents to a particular user with a particular information need. This relates the relevance of a piece of information to a user rather than to a topic. Pertinence is a difficult concept to assess objectively as we have to assume that information needs can be expressed *a priori* in relation to a given database of documents. Also, to assume that needs will not change in relation to what the IR system returns, and that what the user already knows about a particular topic will not affect his assessment of what is pertinent.

Pertinence, however, has a powerful advantage in the development of system, one which motivates the work described here. Although it is difficult for a user to describe what is pertinent, it is easy for her to demonstrate it. Each time a user chooses a document as being of interest he is making a statement about the pertinence of the document and the martial contained within the document, [Kem74] [Cam95]. These statements of pertinence may be used to describe a context within which to describe the user's information needs relative to a collection of information.

In the rest of this paper I aim to describe how some of the features of IR, belief revision and a theory of communication, Relevance Theory, can be to update the system's model of a collection in the light of user's pertinence assessments.

2 RELEVANCE THEORY

The framework proposed here relies heavily on the communication inherent in the user's interaction with the system. The metaphor used here is one of a conversation, the system questions the user by presenting what it regards as germane to the discussion, the most relevant documents. The user answers by selecting the most representative document to her information need.

To realize this dialogue, a formal model of communication is required. This model must be able to cope with the shift in focus as existing information needs and satisfied or as new needs develop in the presence of new evidence.

Two participants in a dialogue may share common terms of reference but it is unlikely that they will share the same connotations for those terms. They both know what they refer to but cannot be sure that the other person is referring to the same object or idea in the same sense or through the same experience. In everyday conversation, we can examine each other's background, experiences, views through the dialogue. This may be acceptable in conversations with another human but it would be tedious for the user of a system to have to explain what she means in detail.

Sperber and Wilson [SW82] address the variety of interpretations available to a speaker and hearer, and use them to develop the Relevance Theory approach. Here what is important is not the knowledge a speaker has about what the hearer knows or believes but the relevance of an utterance of the current context of the conversation.

Uttering a proposition within a context will give rise to contextual implications - non-trivial logical implications that could not be derived solely from the content of the propostion nor from the context in which the proposition was expressed. Both existing information and new evidence combine to infer the contextual implications. Any utterance that produces contextual implications will be relevant to some degree, the degree of relevance being determined by a number of factors, including amount of processing involved to produce the implications and the number of contextual implications generated by the sentence.

Sperber and Wilson encapsulate theory theory in their principle of relevance, from 1982

> **Principle of relevance, (1982)**: speaker tries to express the proposition which is the most relevant one possible to the hearer

Relevance Theory can be described in terms of both classical information retrieval techniques and the system-based dialogue model presented here. Assume that there has been an initial statement of interest from the user, this may be a query or some other representation or description of the user's information need. The initial context is the system's model of the collection that is the terms in the collection and an assessed measure of the utility of that term. The system develops the context by trying to derive contextual implications based upon the user's document selection. In other words it decides how the user's selection affects it's model of what the user is looking for.

This means that we want to use the user's previous areas of interest - the previous contexts - as a basis on which to determine any new contexts. Terms that appear in a user-selected document are counted as being pertinent. This may be over-confident, as a term may not be of interest but only appear due to causal co-occurrence with another term. As we cannot say which terms are of interest we treat all as potentially pertinent.

At this point the work here differs from Relevance Theory as described for human conversation in two ways. Firstly Sperber and Wilson [SW86] talk about a strength of a sentence, this approach insteads uses a probability function distributed over the elements of the discourse - the terms that occur in the collections. Secondly in this model, when the user selects a document she does not only affect the elements of the context but also related terms. That is, although a user is only giving feedback on the information within a document we will view this as feedback on related information that is not present in the document. This is an attempt to prime the system to react to a particular section of information contained within the collection. This we treat as analogous to the contextual implications of Relevance Theory.

3 MODEL

Underpinning this model if a formalism for representing the system's knowledge. The change here is regarded as a form of belief revision. The system's beliefs focus on the terms that appear in the collection, and the fundamental belief that an system has regarding a term t, is "the 'belief that a term is the best term to describe the current stage of the search". The value assigned to the belief $\mathrm{b}(t)$ that states that t is useful in retrieval, will be based upon evidence the user has identified t as being relevant, e.g. through a query, or that there is evidence that t may be relevant, e.g. through co-occurrence of t with a relevant term, or t appearing in a user-selected document. This probabilistic model of belief revision asserts that beliefs of this type can be represented by a probability distribution over the terms space. However it requires that complete knowledge of the structure of the space in terms of the conditional relationships. As such, the revision process, when encountering new information, is less of an expansion/retraction of beliefs but a strengthening/weakening of beliefs based on evidence.

I propose two revisions are necessary, firstly the system must revise its belief system in the light of evidence from the user. Secondly it needs to adapt its new belief system to take into account the relation between its modified beliefs and the evidence that the system possesses about each document. This second revision as shall be shown below does not result in a permanent shift in the system's beliefs about the user's interest.

The first corresponds to the context modification of relevance theory. We are attempting to update the system's view of the user's search by incorporating into the system's belief system new evidence and revising any existing evidence for a belief. For this revision I propose using the form of conditionalisation known as Jeffrey conditionalisation for reasons give in section 3.1.2.

The second revision is different in that it does not result from a change in the system's world. Rather it results from a deliberate attempt to maximise the utility of the system's beliefs in predicting the information most likely to be of interest to the user. It must now try to infer the material in the document from its beliefs. Given a context and a user's response the system is trying to select the interpretations (the documents) that corresponds to what it assesses is the most likely interpretation of the user's utterance. In information retrieval terms what the system is trying to do is assess the documents according to their relevance. The most relevant documents being the most likely interpretations of the system's belief of what the search is for.

For this revision what is required is a general method of revision that can assess the probability of this implication, i.e. $P(b \rightarrow d)$, the probability that a document can be inferred from a set of beliefs. A probability measure is used as Relevance Theory states that only the most likely interpretations are considered. Not all interpretations are equally likely and not all documents are equally relevant to a user. Consequently we must have a way of ranking our interpretations, our measure of a document's relevance, as is also required of an IR system.

A process which fulfills this criterion is known as *logical imaging*, described by Gardenfors [Gar88]. Its application to IR has already been demonstrated and implemented by Crestani and Van Rijsbergen [CvR95, CRSvR95] , and so here I will only attempt to describe the features of imaging necessary to explain the interpretation which I propose using it relative to this model.

3.1 Recasting the Context

Before describing the mechanics of these procedures I want to give some definitions: the **context** is the set of terms in the collection with an associated probability value giving their importance in describing the current search stage, **selected terms** is the set of terms present in the last selected document, **indirectly selected terms** is the set of terms that co-occur with the selected terms, and **unselected terms** is the set of terms that are neither selected nor indirectly selected.

When the user selects a document, the information contained within it should be used to modify the context. The context is revised in the light of evidence from the user. There are three stages in updating the context; first the system's belief in the selected terms should be directly updated. Secondly this change is reflected in the indirectly selected terms. Finally there needs to be a scaling operation to preserve the unit sum of the probabilities.

How this updating on the terms is acheived is threefold depending on which class the term belongs. Selected terms, those that are present in a user-selected document have their probability of importance directly calculated, section 3.1.1. This means that seeing a term in a user-selected document will increase the system's belief in that term's ability to describe the current search stage. Indirectly selected terms, those that have an author-selected relationship will have their probability increased by virtue of being associated with the selected terms. In other words, these terms have their probability increased 'as if it had appeared' in the document. The more selected terms a term appears with the more its probability will be updated 3.1.2. The weakest evidence supplied by the user's selection is for the remaining terms, the unselected ones. If a term does not appear in a user-selected document, nor co-occurs with any terms in that documents, then the user selecting the document does not supply any evidence for the importance of that term. It cannot be said that there is evidence for that term being of no importance, i.e. it probability is zero. What can be said is that there is no additional evidence and so its probability remains unchanged. In fact, as a result of the scaling operation, the probability will be reduced proportional to the changes in the other terms.

3.1.1 Updating

If a term occurs in a set of relevant documents at a level greater than would be expect giving purely statistical relation then we can claim that the term is of importance. The method of how to change the probability of a term should reflect aspects such as the spread of the terms occurrences in the collection, and the frequency within the collection. The actual method chosen is not of primary importance as only scale and relative scale will vary according to the method chosen. When changing the probability we are trying to update our assessment of its discriminatory power to distinguish between relevant and non-relevant documents. A number of methods are being investiagated at the moment including the information radius and the F4 weights, see [Rij79, chapter 6] for an overview of these methods.

3.1.2 Revision

This stage propogates the evidence from updating the selected terms to other terms in the collection. The rationale behind this is that, although we can't directly access the semantic relationships between terms in the mind of the author, or searcher, we can represent the relationships by means of statistical dependencies. If we have no means of directly assessing the probability of importance of a term then it can be determined by using the probability attached to related terms. In theory means propagating belief over all the terms in the term space. In practice we only propagate it to those terms that have an author-selected relationship. This means that I am asserting that the conditional of two terms that do not co-occur is zero, as the probability of co-occurrence is zero.

$$P(term_x \mid term_y) = \frac{P(term_x, term_y)}{P(term_y)} \quad (1)$$

If $term_x$ and $term_y$ do not co-occur then $P(term_x, term_y) = 0$ and consequently $P(term_x|term_y) = 0$. So regardless of individual probability of $term_x$ and $term_y$, seeing $term_y$ alone should allow the assertion that $P(term_x)$ should be zero. Practically, if a term occurs in a document then it will co-occur with every term in that document. This means that $P(term_x)$ will be a factor of the probabilities attached to these terms and the other terms with which it co-occurs.

The system assigns evidence not only according to what is in the collection but but to evidence supplied by the user. The strength of belief in a term is then a measure generated in response to the material selected and to the body of evidence. The beliefs in terms are revised through conditionalisation. The form of conditionalisation I am proposing to use here is Jeffrey conditionalisation. Standard Bayesian conditionalisation gives a means of re-calculating the probability of an event, A, given another event B, The new probability assigned to A, P'(A), is equal to the conditional probability, P(A|B) when P'(B) = 1. Jeffrey conditionalisation generalises this to cases where $P'(B) \neq 1$. This is appropriate here as we are not interested in the probability of events happening, we are interested in the utility of the event, given that it has happened.

Before applying conditionalisation it is necessary to make certain that three conditions hold over the set of terms; exhausivity, mutal exclusivity and stability of the conditional probabilities between terms. The aim is therefore to demonstrate that is possible to calculate the probability of $P(term_x)$ by the equation,

$$P(term_x) = \sum_{n=1}^{N} P(term_x \mid term_n) P(term_n) \quad (2)$$

where N = terms in collection.

Exhausivity:
All the terms in the collection will have an effect on the probability of $term_x$. However the conditional between non-cooccurring terms is zero, (as above) then a large section of the terms will not have an affect on the belief attached to a term. So, even though all terms are considered in theory, in practice only co-occurring terms need be considered for each term.

Mutual exclusivity:
We are measuring the importance of a term. As we have expressed this as the 'belief that a term is the best term to describe the current stage of the search' then we can assume mutual exclusivity of terms in describing this stage.

Given these two conditions hold then we can assert equation 3.

$$P'(term_x) = \sum_{c=1}^{C} P'(term_x \mid term_c) P'(term_c) \quad (3)$$

where C = set of terms that co-occur with $term_x$.

Conditional probabilities do not change:
If we assume the conditional probability reflects the underlying statistical dependency based on co-occurrence then this is valid. This gives,

$$P'(term_x) = \sum_{c=1}^{C} P(term_x \mid term_c) P'(term_c) \quad (4)$$

The new probability of an indirectly selected term depends on the probability of its associated terms, equation 4. These terms belong to two distinct classes; selected and indirectly selected terms. The probability for the selected terms has already been calculated, section 3.1.1. The question is how to the value for each of the indirectly selected terms.

Example: A user selects a document with $term_1$ and $term_2$ in it. The new probability for these terms are calculated directly. For a $term_3$ that co-occurs with terms $term_1$, $term_4$ and $term_5$, the new probability is calculated by the equation,

$$\begin{aligned} P'(term_3) = {} & P(term_3 \mid term_1) P'(term_{t1}) + \\ & P(term_3 \mid term_4) P(term_{t4}) + \\ & P(term_3 \mid term_5) P(term_{t5}) \end{aligned}$$

As there is no direct evidence for the pertinence of $term_4$ and $term_5$, we use the previous values associated to them when calculating the new probability of $term_3$. The probability associated with $term_3$ will increase due to the change in $term_1$.

3.2 Logical Imaging

So far we have calculated the impact of a document selection upon a context, given previous selections.

The next step is to produce a set of system interpretations of the user selection. That is, a list of documents ranked by assessed likelihood of being relevant to the user.

This selection is based on a form of inference, section 2. The system is trying to infer the documents in the collection from the evidence provided. Here there is no query, *per se*. The method used here is to infer the last selected document from the rest of the document collection. More accurately as we are estimating the likelihood of relevance we measure the probability of this inference. Logical imaging revises a probabilistic belief system by moving probability from terms that do not appear in the document to the closest term that does appear. This shift of probability produces a new probability distribution describing the beliefs in relation to the document. The sum of the probabilities attached to those terms appearing in the intersection of the document under consideration and the last user-selected document gives the estimate of relevance for the document.The relative relevance of documents can then be obtained by ranking.

For each new context the amount of probability associated with each term will change. Consequently the effect of changing the context is to alter the relevance of a document is twofold. One, by changing the probability of the term. Secondly by altering the probability of the terms that transfer their probability to the term when imaging.

4 CONCLUSION

Relevance Theory describes how an utterance can give rise to numerous interpretations dependent upon the context. The context and utterance together form a new context, the form of which can only be described by the utterance and particular context. This paper demonstrates how such a model can be described for IR. The context is the set of terms with an associated probability describing its importance in the search. The user selecting a document will cause an increase in the probability of related terms. The overall change preserves the consistency of the set of terms, only the values of the probabilities change. The value assigned to a term depends on the values of its associated terms.

The context is then used, with the last document selected by the user, to produce the new set of relevant documents. As the probabilities change then the measure of similarity should change according to the terms in each document and how they have changed. Also as there is no query modification as such, following [Cam95], the aproach depends only on the user's choice of documents.

Acknowledgements

This work was supported by the FERMI project (8134) - "Formalisation and Expermimentation in the Retrieval of Multimedia Information Retrieval", funded by the European Community under the ESPRIT Basic Research Action.

References

[Cam95] Iain Campbell. Ostensive support for information needs in an adpative information space. In Ian Ruthven, editor, *Miro 95*. Workshops in Computing, Springer Verlag, September 1995.

[CRSvR95] F. Crestani, I. Ruthven, M. Sanderson, and C.J. van Rijsbergen. The troubles with using a logical model of ir on a large collection of documents. Experimenting retrieval by logical imaging on TREC. In *Proceedings of the Fourth Text Retrieval Conference (TREC-4)*, Washington D.C., USA, November 1995.

[CvR95] Fabio Crestani and C J van Rijsbergen. Information retrieval by logical imaging. *Journal of Documentation*, 51(1):3 – 17, March 1995.

[Gar88] Peter Gardenfors. *Knowledge in Flux: Modelling The Dynamics of Epistemic States*. MIT Press, 1988.

[Kem74] D A Kemp. Relevance, pertinence and information systems development. *Information Storage and Retrieval*, 10(2):37 – 47, January 1974.

[Rij79] C.J. van Rijsbergen. *Information Retrieval*. Butterworths, second edition, 1979.

[SW82] Dan Sperber and Deirdre Wilson. Mutual knowledge and relevance in theories of comprehension. In N.V. Smith, editor, *Mutual Knowledge*, pages 61 – 111. Academic Press, 1982. Chapter 2.

[SW86] Dan Sperber and Deirdre Wilson. *Relevance: Communication and Cognition*. Blackwell, 1986.