SENSE RESOLUTION PROPERTIES OF LOGICAL IMAGING

F. CRESTANI
Dipartimento di Elettronica e Informatica
Universitá di Padova, Italy

M. SANDERSON, C.J. VAN RIJSBERGEN
Department of Computing Science
University of Glasgow, Scotland

Abstract

The evaluation of an implication by Imaging is a logical technique developed in the framework of modal logic. Its interpretation in the context of a "possible worlds" semantics is very appealing for IR. In 1994, Crestani and Van Rijsbergen proposed an interpretation of Imaging in the context of IR based on the assumption that "a term is a possible world". This approach enables the exploitation of term–term relationships which are estimated using an information theoretic measure.

Recent analysis of the probability kinematics of Logical Imaging in IR have suggested that this technique has some interesting sense resolution properties. In this paper we will present this new line of research and we will relate it to more classical research into word senses.

1 Introduction

In their recent papers Crestani and Van Rijsbergen [5, 6] described a technique called retrieval by Logical Imaging which originates from the theoretical field of modal logic. They showed how to apply this technique to a text based information retrieval system, and presented a series of experiments that showed Imaging to improve retrieval performance. Crestani and Van Rijsbergen's application of Imaging to IR could be described as a top-down approach: deciding if a technique could be useful to IR from theoretical analysis rather than from the more traditional approach of analysing the results of retrievals.

Once Imaging had been shown to improve retrieval performance, an investigation was undertaken to understand why this technique worked, not in terms of theory but in terms of the "nuts and bolts" of text based IR: words and their meaning. This investigation discovered an unexpected effect that Imaging has on certain types of ambiguous words and it is a description and explanation of this effect that constitutes this paper.

Before describing this effect, an introduction to Imaging is provided followed by a discussion of some pertinent aspects of word sense ambiguity and of automatic disambiguation systems. After this, the effect on word senses caused by Imaging is outlined and this is followed by a description of a small experiment and a proposal for further such experiments. Finally there is a short discussion and conclusions.

2 Logical Imaging and possible worlds semantics

Imaging is a process developed in the framework of Modal Logic [2]. It enables the evaluation of a conditional sentence without explicitly defining the operator "$\rightarrow$". What it requires is a clustering on the space of events (worlds) by means of a primitive relation of neighbourhood. This semantics is called *possible worlds semantic* and it was proposed by Kripke in [11]. According to this semantics the truth value of the conditional $y \rightarrow x$ in a world w is equivalent to the truth value of the consequent x in the closest world w_y where the antecedent y is true. The identification of the closest world is done using the clustering. The passage from a world to another world can be regarded as a beliefs revision, and the passage from a world to its closest is therefore equivalent to the least drastic revision of one's beliefs. Using this process it is possible to implement the *logical uncertainty principle* proposed by Van Rijsbergen in [21]:

> "Given any two sentences x and y; a measure of the uncertainty of $y \rightarrow x$ related to a given data set is determined by the minimal extent to which we have to add information to the data set, to establish the truth of $y \rightarrow x$."

Imaging can be extended to the case where we have a probability distribution on the worlds [13]. A probability distribution over the worlds can be regarded as a measure of the prior uncertainty (or certainty) associated with the beliefs. In this case there is a shift of the original probability P of the world w to the closest world w_y where y is true. Probability is neither created nor destroyed, it is moved from a "not-y-world" to a "y-world" to derive a new probability distribution P_y. This process is called *deriving P_y from P by Imaging on y*.

A formal and detailed exposition of the Imaging process can be found in [19, 13] and we will not report it here. In the next section we will present how Imaging can be used in the context of IR.

3 Retrieving documents by Logical Imaging

Taking into consideration a possible worlds semantics the most obvious way of applying Imaging to IR would be by considering a document as a possible world, regarding it as a set of propositions with associated truth values. This is the view taken originally by an Van Rijsbergen in [22] and followed by others (see [1, 17]). In this view we should evaluate the probability of the conditionals $d \to q$ by computing a new probability distribution P_d by Imaging on d over all the possible worlds, i.e. over all the possible document representations. As we pointed out in [5], there are various problems related to this interpretation of Imaging in IR. Instead, we propose a different approach. We consider the set of terms T, index terms or simply terms used in the document collection, as the set of possible worlds.

In order to apply this approach to Imaging in IR we need a different representation of the document space. We use the artifice of considering a term represented by a set (a vector) of documents. This is the inverse of the representation technique most often used in IR where a document is represented as a set of features, namely terms (or index terms). Intuitively this can be understood as "if you want to know the meaning of a term then look at all the documents in which that term occurs". This idea is not new in IR (see for example [1, 16]) and it has been widely used for the evaluation of term–term similarity. Representing terms in this way, we consider a document d true in a term (world) t if the term t occurs in d, and similarly a query q true in a term (world) t if the term t occurs in it. Using a measure of similarity among terms it is easy to determine the closest term t_d to t that occurs in the document d, or similarly t_q, closest term to t that occurs in the query q.

According to this interpretation of Imaging in IR in [5] we proposed a model called *Retrieval by Logical Imaging* that considers a process of Imaging on d over all the possible terms t in T. This model has been further improved into the *Retrieval by General Logical Imaging* model in [6]. For the purpose of this paper we will refer to the Retrieval by Logical Imaging model, the simplest of the two, which uses Imaging as proposed by Stalnaker in [19]. The properties of Imaging that we will present in this paper with regard to word sense resolution are present in both models.

Retrieval by Logical Imaging is performed by evaluating the following formula:

$$P(d \to q) = P_d(q) = \sum_T P(t)\ I(t_d, q)$$

where

$$I(t_d, q) = \begin{cases} 1 & \text{if } t_d \text{ occurs in } q \\ 0 & \text{otherwise} \end{cases}$$

with t_d as the closest term to t that occurs in d (i.e. where d is true).

This process, called *Imaging on* d, causes a transfer of probabilities from terms not occurring in the document d (i.e. for which the document d is not true) to terms occurring in it (i.e. for which the document d is true).

Similarly we can also evaluate $P(q \to d)$ by *Imaging on* q:

$$P(q \to d) = P_q(d) = \sum_T P(t)\ I(t_q, d)$$

where

$$I(t_q, d) = \begin{cases} 1 & \text{if } t_q \text{ occurs in } d \\ 0 & \text{otherwise} \end{cases}$$

with t_q as the closest term to t that occurs in q (i.e. where q is true).

Here we consider a process of Imaging on q over each possible term t in T so that the probability initially assigned to each term moves from terms not occurring in the query q to terms occurring in the query q.

Retrieval by Logical Imaging can be performed either by Imaging on the document or on the query. Nie showed in [15] that the two conditionals $d \to q$ and $q \to d$ have a very interesting interpretation in the context of IR. The conditional $d \to q$ expresses the *exhaustivity* of the document to a query, i.e. how much of a document content is specified by the query content. In fact $d \to q$ is intuitively equivalent to $q \subseteq d$. The conditional $q \to d$, instead, expresses the *specificity* of a document to a query, i.e. how much of a query content is specified in the document content. In fact, $q \to d$ is intuitively equivalent to $d \subseteq q$. Nie proposed to combine the two measures to produce a "correspondence" measure between query and document. This measure should estimate the relevance of a document to a query.

The application of the above technique to IR requires an appropriate measure of similarity and an appropriate probability distribution over the term space T. In [5] we tackled these problems using a measure of similarity based on a information theoretic measure, the Expected Mutual Information Measure, and a standard IR term weighting technique, the Inverse Document Frequency. In the following sections we will assume as given both a measure of similarity and a probability distribution over the term space.

In the following two sections we explain the two processes of Imaging on the document and on the query by means of an example.

3.1 Evaluation of P(d → q) by Imaging on d

We assume a set of terms T with a probability distribution P which assigns to each term $t \in T$ a probability $P(t)$ so that $\sum P(t) = 1$. We also use the following notation:

$$I(t, x) = \begin{cases} 1 & \text{if } t \text{ occurs in } x \\ 0 & \text{otherwise} \end{cases}$$

We assume we have a document collection D, with $d \in D$, where the documents are represented by terms in the set T. Finally, we assume we have a query q also represented by terms in T. Then, as explained in the previous Section, it is possible to evaluate the $P(d \rightarrow q)$ as:

$$\begin{aligned} P(d \rightarrow q) &= P_d(q) \\ &= \textstyle\sum_T P(t)\ I(t_d, q) \\ &= \textstyle\sum_T P_d(t)\ I(t, q) \end{aligned}$$

where t_d is the term most similar to t which also occurs in d, and $P_d(t)$ is the new probability distribution over the set of terms appearing in d obtained by Imaging on d.

The evaluation of $P(d \rightarrow q) = P_d(q)$ must be repeated for each document in the collection D and it is based on the initial probability distribution over the set of terms T and on the availability of a similarity measure enabling the evaluation of t_d.

For a practical example of this evaluation let us suppose we have a query q described by the terms t_1, t_4, and t_6. We would like to evaluate the probability of relevance of a document d described by terms t_1, t_5, and t_6. Assuming a vector notation, Table 1 reports the evaluation of $P(d \rightarrow q)$ by Imaging on d, that is an estimate of the probability of relevance of the document d to the query q.

The evaluation process is the following:

1. Identify the terms occurring in the document d (third column of the table).
2. Determine for each term in T the t_d, i.e. the most similar term to t for which $I(t, d) = 1$. This is done using the similarity measure on the term space (fourth column).
3. Evaluate $P_d(t)$ by transferring the probabilities from terms not occurring in the document to terms occurring in it (fifth column).
4. Evaluate $t(q)$ for each term, i.e. determine if the term occurs in the query (sixth column).

t	$P(t)$	$I(t,d)$	t_d	$P_d(t)$	$I(t,q)$	$P_d(t) \cdot I(t,q)$
1	0.2	1	1	0.3	1	0.3
2	0.1	0	1	0	0	0
3	0.05	0	5	0	0	0
4	0.2	0	5	0	1	0
5	0.3	1	5	0.55	0	0
6	0.15	1	6	0.15	1	0.15
$\sum_t$	1.0			1.0		0.45

Table 1: Evaluation of $P(d \to q)$ by Imaging on d

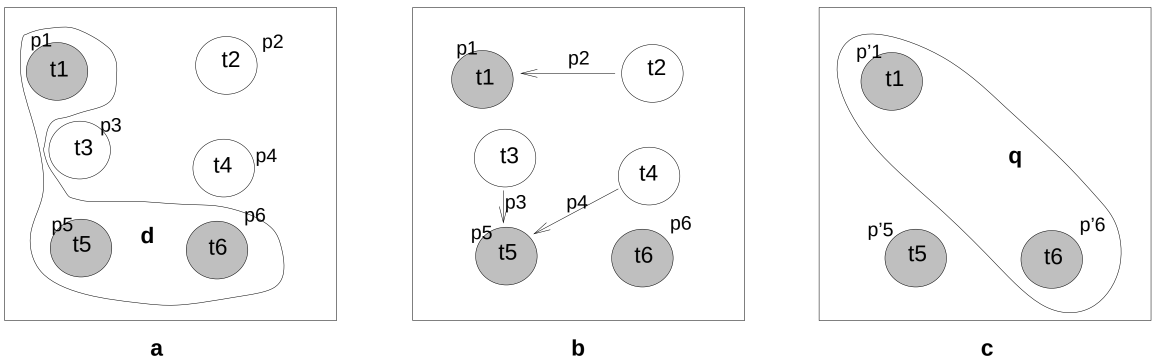


Figure 1: Graphical interpretation of the evaluation of $P(d \to q)$ by Imaging on d.

5. Evaluate $P_d(t)I(t,q)$ for all the terms in the query (seventh column) and evaluate $P_d(q)$ by summation (bottom of seventh column).

Figure 1a shows a graphical representation of this process. As can be seen, each term is represented by a world with its probability measure expressing the importance of the term in the term space T. The shadowed terms occur in document d. We assume a measure of similarity on the term space. Using this information we can now transfer the probability from each term not occurring in the document d to its most similar one occurring in d as depicted in Figure 1b. In Figure 1c the terms with null probability disappear, those occurring in the query q are taken into consideration and their new probabilities $P_d(t)$ are summed up to evaluate $P_d(q)$.

3.2 Evaluation of P(q $\to$ d) by Imaging on q

Using the same data of the previous example we can now evaluate for documents the probability $P(q \to d)$. The terminology is analogous to that of the example above, though modified to take into consideration the evaluation of different elements.

The evaluation of $P(q \rightarrow d)$ is obtained as follows:

$$\begin{aligned} P(q \rightarrow d) &= P_q(d) \\ &= \sum_T P(t)\ I(t_q, d) \\ &= \sum_T P_q(t)\ I(t, d) \end{aligned}$$

where t_q is the term most similar to t which also occurs in q, and $P_q(t)$ is the new probability distribution over the set of terms appearing in q obtained by Imaging on q.

The evaluation of $P(q \rightarrow d)$ must be repeated for each document in the collection D and it is based on the initial probability distribution over the set of terms T and on the availability of a similarity measure enabling the evaluation of t_q.

Table 2 reports an example of the evaluation of $P(q \rightarrow d)$ which can be structured in the following steps:

1. Identify the terms occurring in the query q (third column of the table).
2. Determine for each term in T the t_q, i.e. the most similar term to t for which $I(t, q) = 1$ (fourth column).
3. Evaluate $P_q(t)$ by transferring the probabilities from terms not occurring in the query to terms occurring in it (fifth column).
4. Evaluate $I(t, d)$ for each term, i.e. determine if the term occurs in the document (sixth column).
5. Evaluate $P_q(t) \cdot I(t, d)$ for each term in the document and evaluate $P_q(d)$ by summation (seventh column).

A graphical interpretation of the Imaging process in relation to this example is shown in Figure 2.

4 Word sense ambiguity

Before discussing the relationship between Imaging and word sense ambiguity, a brief overview of some of the features of ambiguity and disambiguation will be presented.

When ever dealing with words, it is important to remember that most words can refer to more than one sense. These individual senses can be quite distinct, for example the

t	$P(t)$	$I(t,q)$	t_q	$P_q(t)$	$I(t,d)$	$P_q(t) \cdot I(t,d)$
1	0.2	1	1	0.35	1	0.35
2	0.1	0	1	0	0	0
3	0.05	0	1	0	0	0
4	0.2	1	4	0.5	0	0
5	0.3	0	4	0	1	0
6	0.15	1	6	0.15	1	0.15
$\sum_t$	1.0			1.0		0.5

Table 2: Evaluation of $P(q \rightarrow d)$ by Imaging on q.

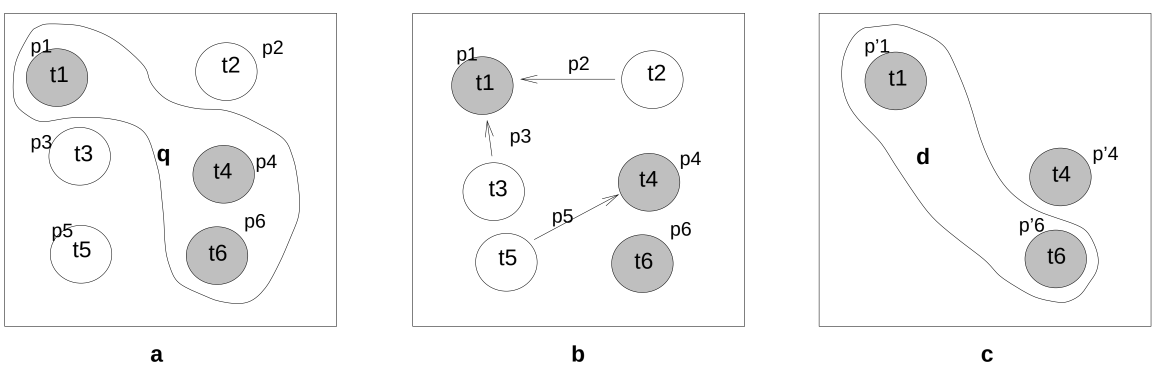


Figure 2: Graphical interpretation of the evaluation of $P(q \rightarrow d)$ by Imaging on q.

word “bat” can refer to an implement used in sports to hit balls or to a furry, flying mammal. Word senses can also be related, for example the word “crash” could refer to a physical event such as a car crash but also it could refer to the shares in a stock market dropping quickly. What sense a word has depends of course on the context that word appears in.

As information retrieval deals with the words of documents, and often ignores their context, inevitably IR is affected by word sense ambiguity. To illustrate, a manager of an on-line news retrieval system reported (in a personal communication with one of the authors) that the current British Prime Minister is causing problems with their retrieval system. A number of users had tried to retrieve articles about the Prime Minister using the query “major”. This query caused many articles about “John Major” to be retrieved. However, in addition many more articles were retrieved where “major” was used as an adjective or as the name of a military rank.

4.1 Word sense disambiguation

The automatic disambiguation (or resolution) of word senses is a problem that has been studied for many years; Gale, Church and Yarowsky [9] cite work dating back to 1950. These disambiguators were used in natural language processing applications such as translation systems. Early attempts to build disambiguators [25, 10, 18] relied on a combination of hand built lexicons and rules. Although working well for the examples they were programmed for, researchers were never able to ’scale up’ the disambiguators to work on large disambiguation problems.

However in the past ten years disambiguation research has moved towards investigating the exploitation of existing corpora already available online. The first work in this area was by Lesk [12] (an often cited paper). He used the textual definitions of an online dictionary to provide evidence for his disambiguator. The use of this evidence can be shown with a simplified example. Suppose we wish to resolve the sense of the word “ash” as it appears in the following sentence.

> *They cleared the* **ash** *from the coal fire.*

To disambiguate “ash”, first its dictionary definition is looked up in the online dictionary and the individual senses of the word are identified. The format of the online dictionary is sufficiently structured to make this identification process relatively simple.

> *ash*: The soft grey powder that remains after something has been burnt.
>
> *ash*: A forest tree common in Britain.

Then the definitions of each of the other words in the sentence (apart from stop words) are looked up as well. For example:

> *coal*: A black mineral which is dug from the earth, which can be burnt to give heat.
>
> *fire*: The condition of burning; flames, light and great heat.

What follows is a process similar to ranked retrieval, where: the individual dictionary sense definitions of "ash" are regarded as a small collection of documents (a collection of two in this case); and the definitions of the words surrounding "ash" are regarded as the query. So the sense definitions are ranked by a scoring function that is based on the number of words co-occuring between a sense's definition and the definitions of each sentence word. The top ranked definition resulting from this process is chosen as the correct sense of "ash".

Lesk performed some limited testing of his technique and reported a disambiguation accuracy of between 50% and 70%. This level of accuracy is actually quite poor as Gale et al [9] found that a disambiguator could have an accuracy of 75% if it always picked the most commonly occurring sense of a word. Although it is likely that if Lesk's disambiguator had incorporated information on the skewed frequency distribution of word senses its performance would have improved. However the importance of Lesk's work was to demonstrate the use of online corpora as sense disambiguation evidence and by doing this, to raise the possibility of building, without to much effort, a disambiguator capable of resolving the senses of a great many words.

Since Lesk's paper a bewildering array of disambiguators have been built using the same principle of collecting sense evidence from a large online corpus and ranking possible word senses according to the degree of match between sense evidence and the context of the ambiguous word: Cowie [4], Wallis [24] and Demetriou [8] have made further use of dictionaries; Zernik [27] built a disambiguator using a morphological analyser; Dagan [7] used bilingual corpora; Church [3] tried aligned bilingual corpora; Voorhees [23] and Sussna [20] used the WordNet thesaurus; and Yarowsky [26] used a combination of Roget's thesaurus and Grollier's encyclopaedia to produce one of the better performing disambiguators to date, achieving 90% accuracy for the 12 words it was tested on.

A shared feature of all the disambiguators referred to above is an assumption that each individual sense of a certain word will appear in a wide context (typically 40–100 surrounding words) that is distinct from the contexts of the other senses of that word. It is not clear if this assumption is entirely correct as research on human disambiguation has found that people can identify word senses accurately from a much narrower context of 1–5 words. This raises the possibility of having two senses of a word occurring in similar wide contexts but in different narrow contexts. Although such a situation probably accounts for some of the errors made by automatic disambiguators, when we consider that the Yarowsky disambiguator (cited above) makes this distinct context assumption and it has a 90% disambiguation accuracy, this assumption appears to be correct most of the time. It is this feature of distinct sense contexts coupled with the skewed frequency distribution of word senses (highlighted by Gale et al) that is important in the relationship between Imaging and the senses of a word.

5 Imaging and sense ambiguity

As has already been discussed, we can have two forms of Imaging in IR: Imaging on the document $P_d(q)$; and Imaging on the query $P_q(d)$. Each form behaves differently with regard to the senses of ambiguous words and will therefore be discussed separately. To illustrate these discussions, a simplified example will be used.

Let us imagine a document collection in which the word "bat" appears in a number of documents and that the frequency of occurrence of its word senses is skewed. In most documents, the word is used to refer to a sporting implement, but occasionally it is used to refer to the flying mammal. As the sporting sense of "bat" is predominant in this collection, words most similar to "bat" (similarity is measured by co-occurrence) will be those similar to this one sense. For this example, let us say that the words most similar to "bat" are "cricket", "baseball", "hit", and "ball". In terms of Imaging, it is these five words that are most likely to transfer their probabilities to each other.

Now let us look at two documents from this collection. Document d_1 is represented by words "bat" and "hit", while document d_2 is represented by words "bat" and "night". Document d_1 uses the word "bat" in the sporting sense (see Figure 3a); document d_2 uses it in the animal sense (see Figure 4a). Suppose a user enters the two word query, "bat", "cricket". How will the two forms of Imaging rank these two documents?

5.1 Imaging on a document

As we recall, when Imaging on a document d, the probabilities of terms not appearing in d are transferred to the terms that do appear in d. The method of transfer is determined by a similarity measure which in this case is approximated using co-occurrence.

Looking at our example, let us first examine d_2. Since the words "cricket", "baseball", "hit", and "ball" are more similar to "bat" than to "night", all their probabilities transfer to this one word (Figure 4b). From Table 4 we can see that this transfer results in document d_2 having an estimated probability of relevance of 0.95.

However in the case of d_1, this document contains the word "hit". As this word is also

t	$P(t)$	$I(t, d_1)$	t_{d_1}	$P_{d_1}(t)$	$I(t, q)$	$P_{d_1}(t) \cdot I(t, q)$
bat	0.2	1	1	0.4	1	0.4
ball	0.1	0	5	0	0	0
night	0.05	0	1	0	0	0
cricket	0.2	0	5	0	1	0
hit	0.3	1	5	0.6	0	0
baseball	0.15	0	1	0	0	0
$\sum_t$	1.0			1.0		0.4

Table 3: Evaluation of $P(d_1 \to q)$ by Imaging on d_1

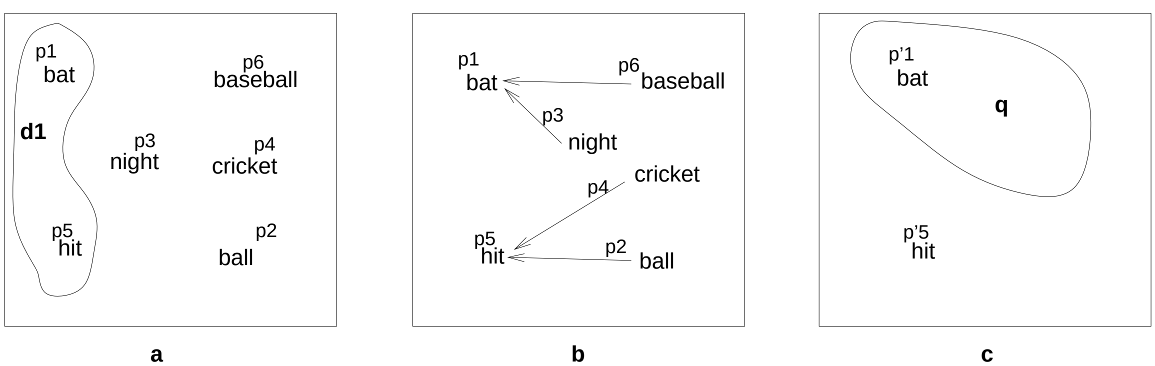
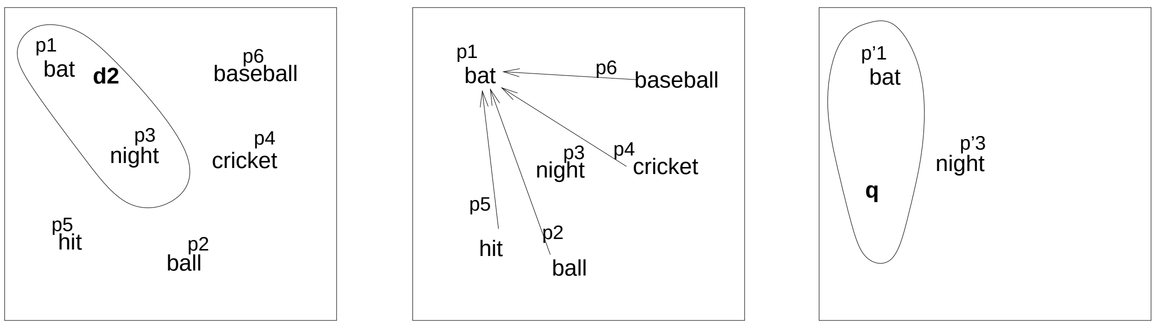


Figure 3: Sense resolution properties of $P(d_1 \to q)$ by Imaging on d_1.

t	$P(t)$	$I(t, d_2)$	t_{d_2}	$P_{d_2}(t)$	$I(t, q)$	$P_{d_2}(t) \cdot I(t, q)$
bat	0.2	1	1	0.95	1	0.95
ball	0.1	0	1	0	0	0
night	0.05	1	3	0.05	0	0
cricket	0.2	0	1	0	1	0
hit	0.3	0	1	0	0	0
baseball	0.15	0	1	0	0	0
$\sum_t$	1.0			1.0		0.95

Table 4: Evaluation of $P(d_2 \rightarrow q)$ by Imaging on d_2.



a b c

Figure 4: Sense resolution properties of $P(d_2 \rightarrow q)$ by Imaging on d_2.

similar to "cricket", "baseball", and "ball", the chances are that the probabilities of some of these words are likely to be transferred to "hit" instead of "bat", this is shown in Figure 3b. As "bat" is the only query word contained in d_1, this results in d_1 having a lower estimated probability of relevance than d_2 (see Table 3), which means that d_2 is ranked higher than d_1!

So what this example seems to show is that Imaging on a document will give preference to those documents which contain query terms appearing in unusual contexts. In terms of word senses, the supposition is that this form of Imaging will rank higher, those documents which hold query terms used in unusual senses.

5.2 Imaging on the query

When Imaging on a query, the method of probability transfer is similar to Imaging on documents except that the transfer is onto the terms in the query. Unlike Imaging on documents this form of Imaging is unaffected by the context in which query terms appear. From Figure 5 it can be seen that the transfer of probabilities to the query

t	$P(t)$	$I(t,q)$	t_q	$P_q(t)$	$I(t,d_1)$	$P_q(t) \cdot I(t,d_1)$
bat	0.2	1	1	0.7	1	0.7
ball	0.1	0	4	0	0	0
night	0.05	0	1	0	0	0
cricket	0.2	1	4	0.3	0	0
hit	0.3	0	1	0	1	0
baseball	0.15	0	1	0.	0	0
$\sum_t$	1.0			1.0		0.7

Table 5: Evaluation of $P(q \to d_1)$ by Imaging on q.

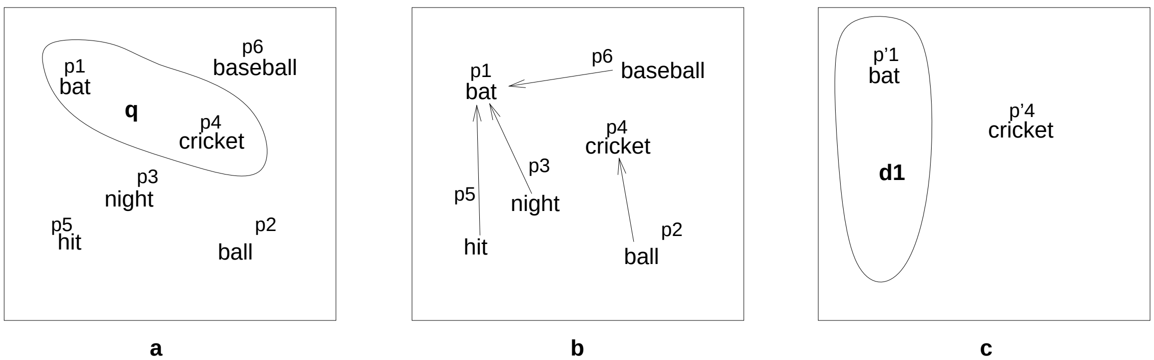


Figure 5: Sense resolution properties of $P(q \to d_1)$ by Imaging on q.

terms is the same regardless of what document is being retrieved. Table 5 shows the estimated probability of relevance for d_1 and it is left as an exercise to the reader to show that d_2 will be assigned the same score.

6 Proposed experimental investigation

While the example given above illustrates the effect discussed here, the extent to which it exists in real retrieval situations and further, the extent to which it improves or degrades retrieval performance needs to be tested. A small test to confirm some of the assumptions made in this paper has been carried out and the results of it are presented here. However it is our intention to proceed with a more complete set of tests and they are described in this section as well.

6.1 Semcor corpus

In a previous section of this paper, Gale et al [9] were quoted as saying that if a disambiguator used a strategy of selecting the most commonly occurring sense, it would be correct 75% of the time. This certainly indicates that the senses of ambiguous words have a skewed distribution. However this figure is calculated from a measure that includes a certain class of ambiguous word that we aren't interested in: namely those words where only one of their senses is used in the corpus. Clearly if we believe that each sense of a word appears in a unique context, then for words of this type, the imaging effect described above would not arise. So we are interested in measuring the frequency distribution of the senses of only those words that have > 1 sense present in a corpus.

It is possible to measure this distribution using the semcor sense tagged corpus[1]. This is a 100,000 word corpus (a subset of the million word Brown corpus) consisting of around 15,000 distinct words, every occurrence of which has been tagged with the word senses defined in the WordNet [14] thesaurus[2].

The following table shows the ratio (expressed as a percentage) of the number of occurrences of a word's most common sense against that word's total number of occurrences, as measured in the semcor corpus. This ratio was computed for separate sets of words, the set a word belongs to is defined by the number of its distinct senses that are present in the corpus. This is different from Gale's measurement which counts the number of senses a word has in a dictionary. Note, a few words in the corpus have > 10 senses, data on these words is not shown in the table.

For each set, the table shows two ratios. The first (the expected ratio) is the value of the ratio that would result if all the word's senses were evenly distributed, this column is put here as a normal to compare the other ratio against. The second ratio is computed from the occurrences of senses within the semcor corpus. As can be seen, a word's most common sense accounts for the majority of that word's occurrences. Over all word sense sets, the most common sense accounts for 61% of occurrences.

This figure is lower than the 75% calculated by Gale (mentioned above) but this is expected because of the omission from the measurement of the type of ambiguous words mentioned above. So from this small experiment we conclude that the senses of a word that are present in a corpus have a skewed distribution for their frequency of occurrence.

[1] No written reference is known for this corpus. At the time this paper was written, the corpus was available at ftp://clarity.princeton.edu/pub.

[2] The sense definitions used in these measurements were those of WordNet 1.4, subsequent versions of WordNet have revised sense definitions and therefore a revised sense tagged semcor corpus. These changes might result in different figures from those presented here, although the differences are anticipated to be small.

$Number\ of$ $word\ senses$	$Expected$ $ratio\,(\%)$	$Ratio\ for\ senses$ $in\ corpus\ (\%)$	$Size$ $of\ set$
2	50	67	2147
3	33	58	767
4	25	53	377
5	20	47	218
6	17	45	130
7	14	43	71
8	13	41	30
9	11	39	20
10	10	33	22

Table 6: The ratio (expressed as a percentage) of the number of occurrences of a word's most common sense against that word's total number of occurrences.

6.2 Proposed retrieval tests

As described in Crestani and Van Rijsbergen [6] experiments have already been carried out that show an improvement in retrieval performance as a result of retrieving document by Logical Imaging. We intend to reexamine the retrieval results of those experiments to determine the extent to which the effects described above are to be found. However, it is believed that these effects may not be so clearly observable when queries with a large number of terms are used. As the queries in the test collections Crestani and Van Rijsbergen used are relatively large, it is intended that further tests be performed on these collections using shorter queries where it is expected that the retrieval results will be more affected by the Imaging-sense effect.

7 Discussion and conclusions

The effect that Imaging on documents has on documents containing ambiguous query terms is caused because the Imaging technique is influenced by all the terms of a document and not just those that appear in the query. It is not clear whether this effect of preferring documents containing query terms in unusual senses or contexts is desirable. Term weighting schemes such as the popular $tf \cdot idf$ do give preference to unusual terms appearing in a document in unusually large quantities. Therefore one might think that this preference for the unusual might indicate that the Imaging effect is desirable. However if a user enters a query term it would seem reasonable to expect him to intend the most common sense. Until the tests outlined above are completed though, we prefer to withhold our judgement.

References

[1] G. Amati and S. Kerpedjiev. An Information Retrieval logical model: implementation and experiments. Technical Report Rel 5B04892, Fondazione Ugo Bordoni, Roma, Italy, March 1992.

[2] B.F. Chellas. *Modal logic: an introduction*. Cambridge University Press, Cambridge, UK, 1980.

[3] K.W. Church. Using bilingual materials to develop word sense disambiguation methods. In *Proceedings of ACM SIGIR*, page 350, Copenhagen, Denmark, June 1992.

[4] J. Cowie, J. Guthrie, and L. Guthrie. Lexical disambiguation using simulated annealing. In *Proceedings of the COLING Conference*, pages 359–365, August 1992.

[5] F. Crestani and C.J. van Rijsbergen. Information Retrieval by Logical Imaging. *Journal of Documentation*, 51(1):1–15, 1995.

[6] F. Crestani and C.J. van Rijsbergen. Probability kinematics in information retrieval: a case study. Technical Report FERMI/95/1, ESPRIT Basic Research Action, Project Number 8134 - FERMI, Department of Computing Science, Glasgow University, January 1995.

[7] I. Dagan, A. Itai, and U. Schwall. Two languages are more informative than one. In *Proceedings of the ACL*, pages 130–137, 1991.

[8] G.C. Demetriou. Lexical disambiguation using constraint handling in prolog (chip). In *Proceedings of the European Chapter of the ACL*, pages 431–436, 1993.

[9] W. Gale, K.W. Church, and D. Yarowsky. Estimating upper and lower bounds on the performance of word-sense disambiguation programs. In *Proceedings of the ACL*, pages 249–256, 1992.

[10] E. Kelly and P. Stone. *Computer recognition of english word senses*. North-Holland Publishing Co, Amsterdam, 1975.

[11] S.A. Kripke. Semantical considerations on modal logic. In L. Linsky, editor, *Reference and modality*, chapter 5, pages 63–73. Oxford University Press, Oxford, UK, 1971.

[12] M. Lesk. Automatic sense disambiguation: how to tell a pine cone from an ice cream cone. In *Proceedings of the SIGDOC Conference*, pages 24–26, June 1986.

[13] D. Lewis. Probability of conditionals and conditionals probabilities. In W.L. Harper, R. Stalnaker, and G. Pearce, editors, *Ifs*, The University of Western Ontario Series in Philosophy of Science, pages 129–147. D.Reidel Publishing Company, Dordrecht, Holland, 1981.

[14] G. Miller. Special issue, WordNet: An on-line lexical database. *International Journal of Lexicography*, 4, 1990.

[15] J.Y. Nie. An outline of a general model for information retrieval. In *Proceedings of ACM SIGIR*, pages 495–506, Grenoble, France, June 1988.

[16] Y. Qiu and H.P. Frei. Concept based query expansion. In *Proceedings of ACM SIGIR*, pages 160–171, Pittsburgh, PA, USA, June 1993.

[17] T.M.T Sembok and C.J. van Rijsbergen. Imaging: a relevance feedback retrieval with nearest neighbour clusters. In *Proceedings of the BCS Colloquium in Information Retrieval*, pages 91–107, Glasgow, UK, March 1993.

[18] S. Small and C. Rieger. *Strategies for Natural Language Processing*. LEA, 1982.

[19] R. Stalnaker. Probability and conditionals. In W.L. Harper, R. Stalnaker, and G. Pearce, editors, *Ifs*, The University of Western Ontario Series in Philosophy of Science, pages 107–128. D.Riedel Publishing Company, Dordrecht, Holland, 1981.

[20] M. Sussna. Word sense disambiguation for free-text indexing using a massive semantic network. In *Proceedings of CIKM*, 1993.

[21] C.J. van Rijsbergen. A non-classical logic for Information Retrieval. *The Computer Journal*, 29(6):481–485, 1986.

[22] C.J. van Rijsbergen. Toward a new information logic. In *Proceedings of ACM SIGIR*, Cambridge, USA, June 1989.

[23] E.M. Voorhees. On expanding query vectors with lexically related words. In *Proceeding of the TREC Conference*, pages 223–232, Gaithersburg, Maryland, USA, August 1993.

[24] P. Wallis. Information retrieval based on paraphrase. In *Proceedings of PACLING*, 1993.

[25] S.F. Weiss. Learning to disambiguate. *Information Storage and Retrieval*, 9:33–41, 1973.

[26] D. Yarowsky. Word sense disambiguation using statistical models of Roget's categories trained on large corpora. In *Proceedings of the COLING Conference*, pages 454–460, August 1992.

[27] U. Zernik. TRAIN1 vs. TRAIN2: agging word senses in corpus. In *Proceedings of RIAO, Intelligent Text and Image Handling*, pages 567–585, 1991.