Analysis of the effect of negation on information retrieval of medical data

Bevan Koopman[*]
Australian e-Health Research Centre
CSIRO
QLD 4029 Australia
bevan.koopman@csiro.au

Peter Bruza
Faculty of Science & Technology
Queensland University of Technology
QLD 4001 Australia
p.bruza@qut.edu.au

Laurianne Sitbon
Faculty of Science & Technology
Queensland University of Technology
QLD 4001 Australia
laurianne.sitbon@qut.edu.au

Michael Lawley
Australian e-Health Research Centre
CSIRO
QLD 4029 Australia
michael.lawley@csiro.au

Abstract *Most information retrieval (IR) models treat the presence of a term within a document as an indication that the document is somehow "about" that term, they do not take into account when a term might be explicitly negated. Medical data, by its nature, contains a high frequency of negated terms – e.g. "review of systems showed no chest pain or shortness of breath".*

This papers presents a study of the effects of negation on information retrieval. We present a number of experiments to determine whether negation has a significant negative effect on IR performance and whether language models that take negation into account might improve performance. We use a collection of real medical records as our test corpus. Our findings are that negation has some effect on system performance, but this will likely be confined to domains such as medical data where negation is prevalent.

Keywords Information Retrieval, Natural Language Techniques and Documents

1 Introduction

Consider the extract below taken from a patient's medical record:

> "Review of systems is significant for subjective chills and fever with a temperature of 104 this morning. Review of systems is otherwise **negative** for headache, chest pain, shortness of breath, dysuria, or increased frequency of urination." [7, #22248]

Most information retrieval systems would consider queries for "headache" and "chest pain" as good matches for the above document, the assumption being that the presence of a term denotes relevance. For documents that contain little or no negation this may not pose any significant problem, but medical data by its nature contains a high degree of explicit negation [8]. This begs the question of what effect does the prevalence of negation in medical data have on medical information retrieval. Averbuch et al. estimate that ignoring negations in medical narrative reports can reduce retrieval performance by as much as 40% [1].

In this paper we present a number of empirical studies on the effect of negation on current state-of-the-art IR systems. Our test corpus is a collection of medical records and test queries are commonly negated medical terms.

2 Related work

This section summarises some of the work to date on dealing with negation in information retrieval related fields. Much of the focus on negation is in the computational linguistics and NLP fields, less work has focused on negation in retrieval tasks.

This study focuses on explicitly negated terms found in documents and differs from other work concerned with negation in queries, for example the Boolean query "spider AND web NOT internet". Dealing with negation in queries presents its own set of problems, as outlined by McQuire & Eastman [6]. The solution is to exclude documents containing the negated term from the result set. There have been a number techniques to achieve this, these include: post-retrieval filtering [5], negative-scoring, negative-relevance feedback [2] and vector negation [9]. All these approaches focus on negation in the query and do not consider negated terms found in a document.

Prior work for dealing with negation in documents has primarily been done within the Natural Language Processing (NLP) community. The main focus here is on negation detection or recognition – analysing the

---

[*]Also Faculty of Science & Technology, Queensland University of Technology

Proceedings of the 15th Australasian Document Computing Symposium, Melbourne, Australia, 10 December 2010.

syntax of natural language to determine which terms have a negative context. A difficult problem is determining the scope of negated terms when negation is detected [4]. This can even prove difficult for human subjects [6]. Many of the solutions to negation detection have been within the application area of dealing with medical data [8, 1], a reflection of the prevalence and importance of negation in medical narratives. NegEx is one popular open source tool for identify negated terms in clinical texts [3]. All these solutions are concerned only with negation detection, they do not propose methods for dealing with negation in the next step of information retrieval.

This paper intends to consider what happens after negation detection. We first provide an empirical analysis of the effect of negation on information retrieval tasks. This is intended to provide the motivation for whether further work on a unified method for negation in IR is justified.

3 Methods

This section provides details of three separate experiments we undertook to investigate the effects of negation on a corpus of medical records.

As our baseline IR system we use the Indri search engine[1] with Porter stemmer for indexing and BM25 term weighting for retrieval. A small comparison of Indri with Lucene showed similar results.

As our test corpus we use the BLULab NLP repository [7], a collection of 81,617 de-identified clinical reports from multiple U.S. hospitals during 2007.

3.1 Experiment A – common negated medical terms

This initial experiment aimed to identify commonly negated terms from the BLULab medical corpus. This was implemented by searching the corpus for the single term appearing after the negation qualifiers: "no", "negative", "negative for" and "not". The number of occurrences matching this pattern for each term was recorded. Terms were then ranked in descending order of the number of negation occurrences.

3.2 Experiment B – precision@10 for negated terms

From the commonly negated terms identified in Experiment A the top 15 (stemmed) terms representing common medical concepts were chosen as candidate queries. These were: *murmur*, *fever*, *fractur*, *edema*, *rash*, *jvd*, *pneumothorax*, *nausea*, *smoke*, *lymphadenopathi*, *mass*, *club*, *wheez*, *headach* and *cyanosi*.

These queries were submitted to the Indri baseline IR system and the top 10 results analysed for their relevance, this gave a measure of precision@10.

[1] http://www.lemurproject.org/indri

3.3 Experiment C – relevance ratio for entire results list

This experiment looked further than precision @ 10 by analysing the entire result set rather than just the top 10 results. The same queries were used as Experiment B (*murmur*, *fever*, etc.). For each query the entire retrieval list was analysed to determine what portion of documents contained the term in negative form and the term in positive form. This gave a relevance ratio for each query q, this is calculated as:

$$\mathrm{rel}(q) = \frac{\text{documents without negation}}{\text{total matching documents}}$$

The experiment was repeated using the top 200 (rather than top 15) negated terms.

A document that contains the term in both positive and negative form would appear in both the lists of positive and negative occurrences for that term.

4 Results

Results of the three experiments are presented in the following subsections. The analysis and interpretation of the results is provided separately in the Discussion, Section 5.

4.1 Experiment A – common negated medical terms

Table 1 presents terms from the BLULab medical corpus that are commonly found in negated form. Terms are ordered in descending frequency of negation occurrences. The terms highlighted in **bold** are the top medical terms chosen as queries for subsequent experiments.

Term	Occurrences	Term	Occurrences
evid	19626	**nausea**	3256
acut	19455	abdomin	3122
have	7951	**smoke**	3115
signific	7856	**lymphadenopathi**	2964
for	7809	had	2883
murmur	6665	short	2793
known	6527	**mass**	2714
other	5722	show	2636
chest	5438	appar	2634
focal	5139	appear	2558
fever	4878	**club**	2510
chang	4690	obvious	2506
fractur	4451	been	2422
histori	4376	activ	2359
edema	4011	**wheez**	2313
be	3953	**headach**	2309
rash	3769	free	2233
jvd	3676	**cyanosi**	2137
definit	3524	abnorm	2035
pneumothorax	3297	prior	2026

Table 1: Commonly negated terms from medical records. Terms in **bold** were chosen as queries.

4.2 Experiment B – precision@10 for negated terms

Table 2 presents precision measures for the top 10 documents returned by each of the 15 queries of commonly negated medical terms. Figure 1 presents these results graphically.

Term	**Prec@10**	**Term**	**Prec@10**
murmur	1.0000	smoke	0.9000
fever	0.9000	lymphadenopathi	0.8000
fractur	0.5000	mass	0.9000
edema	0.9000	club	0.3000
rash	0.8000	wheez	1.0000
jvd	0.3000	headach	1.0000
pneumothorax	0.9000	cyanosi	0.7000
nausea	1.0000		
Average	**0.86**		

Table 2: Precision for top 10 ranked documents for commonly negated medical terms.

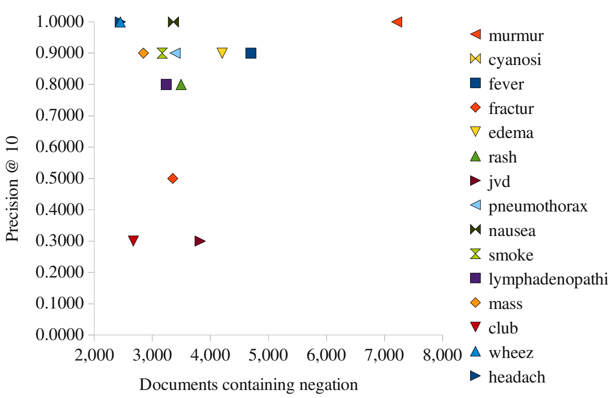

Figure 1: Correlation between negation occurrence and precision @ 10.

4.3 Experiment C – relevance ratio

This experiment presents the relevance ratio – what portion of the entire result set for each query contains the term in positive form. The experiment was done twice, once for the top 15 negated terms and once for the top 200 terms.

Table 3 presents the results for the relevance ratio for the top 15 negated terms. These results are presented in graphical form in Figure 2.

The second part of the experiment was to determine the relevance ratio for the top 200 negated terms, results represented in Figure 3.

Query	**Total**	**Documents with negation**	**Relevance ratio**
murmur	13,573	7,210	0.4688
fever	16,862	4,699	0.7213
fractur	14,194	3,353	0.7638
edema	24,582	4,204	0.8290
rash	7,278	3,495	0.5198
jvd	5,075	3,825	0.2463
pneumothorax	8,428	5,035	0.5974
nausea	15,417	3,365	0.7817
smoke	10,940	3,169	0.7103
lymphadenopathi	7,093	3,241	0.5431
mass	13,569	2,846	0.7903
club	5,823	2,673	0.5410
wheez	6,744	2,448	0.6370
headach	9,322	2,449 3	0.7373
cyanosi	6,649	2,201	0.6690
Average	**11,037**	**3,505**	**0.6371**

Table 3: Relevance ratio – what portion of the entire result set for each query contains the term in negated form.

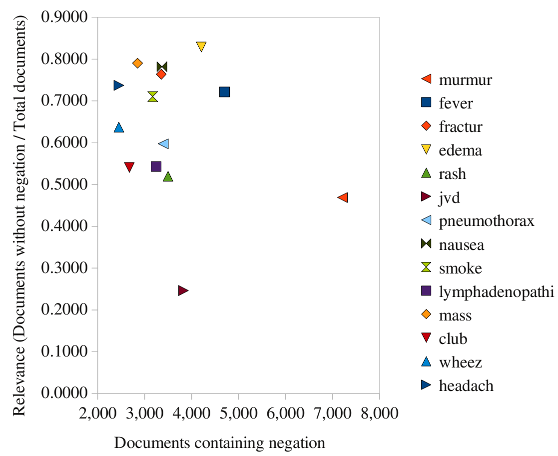

Figure 2: Correlation between negation occurrence and relevance ratio. Top 15 terms.

5 Discussion

The results from Experiment B (see Section 4.2) present the precision @ 10 measurement. Overall the baseline system performs well with an average precision of 0.86. In most cases documents containing the negated form were not found in the top 10 results. The reason for this is that when a term occurs in negated form it typically only occurs once within the document, for example a document will have a single mention of "no rash". In contrast when the term appears in positive form it typically appears a number of times – a medical record relating to someone suffering from a rash will mention the term "rash" multiple times. The standard termweighting function will rank the document containing multiple positive occurrences of "rash" above that of the single negative occurrence. In this way current IR systems implicitly deal with negation by their standard document / term frequency weighting functions.

In Experiment C we considered the entire result set returned (rather than just the top 10 documents). Here negation had a more marked affect, average precision was 0.6371. However, there was no strong correlation between the occurrence of negation and performance (as shown in Figure 2). "JVD" and "murmur" were two queries that performed well below the average, these two terms are part of a standard observation doctors per-

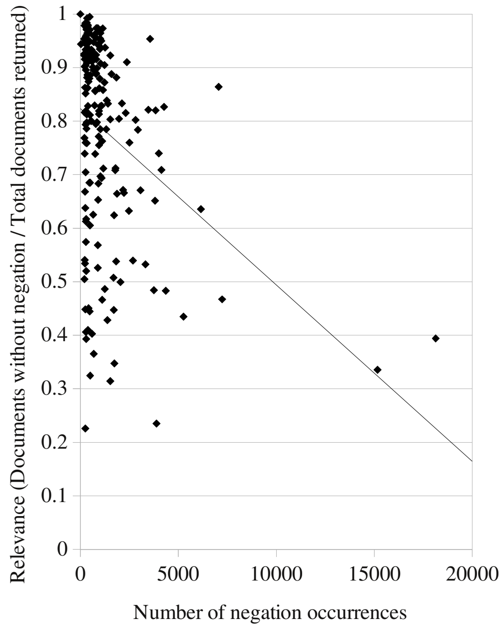

Figure 3: Correlation between negation occurrence and relevance ratio. Top 200 terms.

form on all patients and therefore nearly always appear in a patient's record in negated form.

Overall negation does not have a major impact on retrieval. Term weighting functions are effective at ranking documents with negated terms. We conclude that specific methods of dealing with negation would only be required for specific domains such as medical data where negation is prevalent and may pose problems in the quality of results retrieved.

5.1 Limitations & future work

In our experiments negation detection was implemented by searching the corpus for the single term appearing after the negation qualifiers: "no", "negative", "negation for" and "not". This simplistic approach would not identify more complex examples such as "history inconsistent with stroke" or "patient denies any pain". Additionally we do not identify negated conjunctions like the example presented in the introduction – "…negative for headache, chest pain, shortness of breath, …". We would only identify "headache" as a negated term from this extract.

Implementing a best-practise NLP negation detection tool (e.g. NegEx) would likely increase the negative effects of negation on the relevance ratio results (Experiment C). It is, however, unlikely to affect the precision @ 10 results, which we believe is the more important indicator.

6 Conclusion

We have presented medical data as a domain where negation in documents is prevalent. Based on this we have conducted a number of experiments to determine what effect the high prevalence of negation has on information retrieval. The purpose of which is to determine whether specific methods of dealing with negation might be developed to improve retrieval performance. Our findings are that modern term-weighting functions used in IR systems are quite effective at dealing with negation and that specific methods for dealing with negation are only really relevant to specific domains such as dealing with medical data.

7 Acknowledgements

This study was supported by the Australian e-Health Research Centre. We thank the University of Pittsburgh BLULab group [7] for providing the test corpus of medical records.

References

[1] Mordechai Averbuch, Tom H. Karson, Benjamin Ben-Ami, Oded Maimond and Lior Rokachd. Context-sensitive medical information retrieval. In *Proceedings of the 11th World Congress on Medical Informatics (MEDINFO-2004)*, San Francisco, USA, 2004.

[2] Mark D. Dunlop. The effect of accessing non-matching documents on relevance feedback. *ACM Transactions on Information Systems*, Volume 15, Number 2, pages 137 – 153, 1997.

[3] Ilya M. Goldin and Wendy W. Chapman. Learning to detect negation with 'not' in medical texts. In *Workshop at the 26th ACM SIGIR Conference*, Toronto, Canada, 2003.

[4] Laurence R. Horn. *A natural history of negation*. University of Chicago Press, Chicago, IL, 1989.

[5] Gerard Saltonand Michael J. McGill. *Introduction to modern information retrieval*. McGraw-Hill Book Company, New York, 1984.

[6] April R. McQuire and Caroline M. Eastman. The ambiguity of negation in natural language queries to information retrieval systems. *Journal of the American Society for Information Science*, Volume 49, Number 8, pages 686 – 692, 1998.

[7] University of Pittsburgh. BLULab NLP Repository. http://nlp.dbmi.pitt.edu/nlprepository.html, July 2010.

[8] Lior Rokach, Roni Romano and Oded Maimon. Negation recognition in medical narrative reports. *Information Retrieval*, Volume 11, pages 499 – 538, 2008.

[9] Dominic Widdows. Orthogonal negation in vector spaces for modelling word-meanings and document retrieval. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics*, Morristown, USA, 2003.