Orthogonal negation for document re-ranking[*]

Pierpaolo Basile, Annalina Caputo, and Giovanni Semeraro

Dept. of Computer Science - University of Bari "Aldo Moro"
Via Orabona, 4 - I-70125, Bari (ITALY)
basilepp@di.uniba.it, acaputo@di.uniba.it, semeraro@di.uniba.it

Abstract. In this work, we propose a method for document re-ranking, which exploits negative feedback represented by non-relevant documents. The concept of non-relevance is modelled through the quantum negation operator. The evaluation carried out on a standard collection shows the effectiveness of the proposed method in both the classical Vector Space Model and a Semantic Document Space.

1 Introduction

This work investigates the role of non-relevant documents in document re-ranking. Classic relevance feedback methods are able to handle negative feedback by subtracting "information" from the original query. However, these approaches suffer from the side effect caused by information loss. To deal with this effect, we propose a negative feedback based on quantum negation that is able to remove only the unwanted aspects pertaining to non-relevant documents. The key idea behind our approach is to build a document vector d^* corresponding to an *ideal document* which best fits the user's need, and then re-rank the initial set of ranked documents D_{init} by computing the similarity between d^* and each document in D_{init}. The ideal document vector d^* should fit the *concepts* in the set of relevant documents D^+, while skipping *concepts* in the set D^- of non-relevant ones. Formally, a new relevance score is computed for each document $d_i \in D_{init}$ according to the following equation:

$$S(d_i) = \alpha * S_{D_{init}}(d_i) + (1 - \alpha) * sim(d_i, d^*) \tag{1}$$

where $S_{D_{init}}(d_i)$ is the score of d_i in the initial rank D_{init}, while $sim(d_i, d^*)$ is the similarity degree between the document vector d_i and the ideal document vector d^* computed by cosine similarity. The outcome of the process is a list of documents ranked according to the new scores computed using Equation 1. In our approach, documents are represented as vectors in a geometric space in which similar documents are represented close to each other. This space can be the classical *Vector Space Model* (VSM) or a *Semantic Document Space* (SDS)

[*] This paper summarizes the main results already published in Basile, P., Caputo, A., Semeraro, G.: Negation for document re-ranking in ad-hoc retrieval. In: Amati, G., Crestani, F. (eds.) Advances in Information Retrieval Theory, Lecture Notes in Computer Science, vol. 6931, pp. 285–296. Springer Berlin / Heidelberg (2011)

induced by a distributional approach. Moreover, we compare our strategy with a classical strategy based on "information subtraction".

2 Re-ranking using quantum negation

To build the ideal document d^* we use a geometrical space where d^* is computed as a vector close to relevant documents and unrelated to non-relevant ones. In our space the concept of relevance is expressed in terms of similarity, while the concept of irrelevance is defined by orthogonality (similarity equals to zero). Formally, we want to compute the vector which represents the following logical operation:

$$d^* = d_1^+ \vee d_2^+ \vee \ldots \vee d_n^+ \wedge NOT(d_1^-) \wedge NOT(d_2^-) \wedge \ldots \wedge NOT(d_m^-) \tag{2}$$

where $D^+ = \{d_i^+, i = 1 \ldots n\}$ and $D^- = \{d_j^-, j = 1 \ldots m\}$ are the subsets of relevant and non-relevant documents respectively.

As shown in [5], given two vectors a and b in a vector space V endowed with a scalar product, $a \quad NOT \quad b$ corresponds to the projection of a onto the orthogonal space $\langle b \rangle^\perp \equiv \{v \in V : \forall b \in \langle b \rangle, v \cdot b = 0\}$, where $\langle b \rangle$ is the subspace $\{\lambda b : \lambda \in \mathbb{R}\}$. Equation 2 consists in computing a vector which represents the disjunction of the documents in D^+, and then projecting this vector onto all m orthogonal spaces defined by the documents in D^-. This operation is quite complex to compute, but applying De Morgan rules to the conjunction of negations, it can be transformed in a single negation of disjunctions:

$$d^* = d_1^+ \vee d_2^+ \vee \ldots \vee d_n^+ \wedge NOT(d_1^- \vee d_2^- \vee \ldots \vee d_m^-) \tag{3}$$

Thus, it is possible to build the ideal document vector d^* in two steps:

1. compute the disjunction of relevant documents as the vector sum of relevant documents. Indeed, disjunction in set theory is modelled as set union, which corresponds to the vector sum in linear algebra;
2. compute the projection of the vector sum of relevant documents onto the orthogonal space defined by the vector sum of non-relevant documents, for example using the Gram-Schmidt method. This means that the result vector captures those aspects that are common to relevant documents and are distant from non-relevant ones.

Disjunction and negation using quantum logic are thoroughly described in [5]. An overview of Quantum Mechanics for Information Retrieval can be found in [2]. Finally, the re-ranking algorithm is performed by computing the Equation 1.

3 Evaluation and Remarks

The aim of our evaluation is twofold. We want to prove that our re-ranking strategy based on quantum negation improves retrieval performance and outperforms the "information subtraction" method. To perform re-ranking using a classical "information subtraction" strategy, we assume that documents are represented by classical bag-of-words. Given D^+ and D^-, the computation of the ideal document d^*_C is based on the Rocchio [4] algorithm as follows:

$$d^*_C = \frac{1}{|D^+|} \sum_{i \in D^+} d_i - \frac{1}{|D^-|} \sum_{j \in D^-} d_j \qquad (4)$$

Moreover, we want to evaluate the performance of our approach when a reduced space, likewise a *Semantic Document Space*, is involved. The SDS is built by *Random Indexing* (RI) [1] a technique based on the Random Projection: the idea is that high dimensional vectors chosen randomly are "nearly orthogonal". This yields a result that is comparable to orthogonalization methods, such as Singular-Value Decomposition, but saving computational resources.

We set up a baseline system based on the BM25 multi-fields model [3].

The evaluation has been designed using the CLEF 2009 Ad-Hoc WSD Robust Task collection. To evaluate the performance we performed 150 runs by considering all possible combinations of the three parameters involved in our method: n (the cardinality of D^+), m (the cardinality of D^-) and the parameter α used for the linear combination of the scores (see Equation 1). We selected different ranges for each parameter: n ranges in $[1, 5, 10, 20, 40]$, m in $[0, 1, 5, 10, 20, 40]$, while α in $[0.3, 0.4, 0.5, 0.6, 0.7]$. The cardinality of D_{init} was set to 1,000.

Identifying relevant documents is quite straightforward: we assume the top ranked documents as relevant, while identifying non-relevant ones is not trivial. We proposed two strategies to select the set (D^-) of non-relevant documents, which are based on plausible heuristics rather than a theory:

1. *BOTTOM*, which selects the non-relevant documents from the bottom of the rank;
2. *RELJUD*, which relies on relevance judgements provided by CLEF organizers. This technique selects the top m ranked documents which are non-relevant exploiting the relevance judgements. We use this strategy to "simulate" the user's explicit feedback; in other words we assume that the user selects the first m non-relevant documents.

We evaluate each run in terms of MAP and GMAP over all the queries. Table 1 reports the results for the *baseline* and all three strategies (*Information Subtraction*, *VSM* and *SDS*). For each strategy, *positive* stands for the best run when only relevant documents were involved, while *BOTTOM* and *RELJUD* indicate the best run obtained for both strategies respectively. Improvements in percentage ($\Delta\%$) with respect to the baseline are reported.

The experimental results are very encouraging. Both methods (*BOTTOM* and *RELJUD*) show improvements with respect to the baseline in all the approaches. The main outcome is that quantum negation outperforms the "information subtraction" strategy.

Genarally, *BOTTOM* strategy results in not significant improvements, and in the case of "information subtraction", the introduction of non-relevant documents results in lower performance. The blind selection of non-relevant documents produces a side effect in "information subtraction" strategy due to the

Table 1. Evaluation results using all three strategies.

Method	*Run*	n	m	α	*MAP*	*Δ%*	*GMAP*	*Δ%*
-	baseline	-	-	-	0.4139	-	0.1846	-
Information Subtraction	positive	1	0	0.6	0.4208	+1.67	0.1754	-4.98
	BOTTOM	1	1	0.6	0.4175	+0.87	0.1750	-5.20
	RELJUD	40	40	0.7	0.5932	+43.32	0.2948	+59.70
Orthogonalization VSM	positive	1	0	0.5	0.4372	+5.63	0.1923	+4.17
	BOTTOM	1	5	0.6	0.4384	+5.92	0.1923	+4.17
	RELJUD	40	40	0.7	0.6649	+60.64	0.3240	+75.51
Orthogonalization SDS	positive	1	0	0.5	0.4362	+5.39	0.1931	+4.60
	BOTTOM	1	5	0.6	0.4367	+5.51	0.1928	+4.44
	RELJUD	40	40	0.7	0.6646	+60.57	0.3415	+84.99

information loss, while the quantum negation removes from relevant documents only those "negative" aspects that belong to the non-relevant ones.

As expected, the method *RELJUD* obtains very high results. In this case quantum negation obtains very high improvements with respect to the "information subtraction" strategy. This proves that quantum negation is able to take advantage of information about non-relevant documents. The best results in *RELJUD* are obtained when a lot of non-relevant documents are involved, but in a real scenario this is highly improbable. We performed several runs considering only one non-relevant document and varying the numbers of those relevant. The highest MAP value for *SDS* is 0.4606 (GMAP=0.2056), while for *VSM* is 0.4588 (GMAP=0.2028), both values are obtained with five relevant documents (these results are not reported for the sake of simplicity). Moreover, in both *BOTTOM* and *RELJUD* differences between *SDS* and *VSM* are not relevant.

These values support our thesis that negation expressed by quantum logic operator is able to model effectively the concept of non-relevance, opening new perspective for those tasks where the concept of non relevance plays a key role.

References

1. Kanerva, P.: Sparse Distributed Memory. MIT Press (1988)
2. Melucci, M., Rijsbergen, K.: Quantum mechanics and information retrieval. In: Melucci, M., Baeza-Yates, R. (eds.) Advanced Topics in Information Retrieval, Information Retrieval, vol. 33, pp. 125–155. Springer (2011)
3. Robertson, S., Zaragoza, H., Taylor, M.: Simple BM25 extension to multiple weighted fields. In: Proc. of the 13th ACM Int. Conf. on Information and Knowledge Management. pp. 42–49. ACM, New York, NY, USA (2004)
4. Salton, G., Buckley, C.: Improving retrieval performance by relevance feedback. Journal of the American Society for Information Science 41(4), 288–297 (1990)
5. Widdows, D., Peters, S.: Word vectors and quantum logic: Experiments with negation and disjunction. Mathematics of language (8), 141–154 (2003)