A Model for Information Retrieval based on Possibilistic Networks

Asma H. BRINI and Mohand BOUGHANEM and Didier DUBOIS

IRIT
118, route de Narbonne, Cedex 4
Toulouse, FRANCE

Abstract. This paper proposes a model for Information Retrieval (IR) based on possibilistic directed networks. Relations documents-terms and query-terms are modeled through possibility and necessity measures rather than a probability measure. The relevance value for the document given the query is measured by two degrees: the necessity and the possibility. More precisely, the user's query triggers a propagation process to retrieve necessarily or at least possibly relevant documents. The possibility degree is convenient to filter documents out from the response (retrieved documents) and the necessity degree is useful for document relevance confirmation. Separating these notions may account for the imprecision pervading the retrieval process. Moreover, an improved weighting of terms in a query not present in the document is introduced. Experiments carried out on a sub-collection of CLEF, namely LeMonde 1994, a French newspapers collection, showed the effectiveness of the model.

1 Introduction

The Information Retrieval (IR) process consists in selecting among a large collection a set of documents that are relevant to a user's query. The set of retrieved documents in answer to a query does not usually correspond to the set of documents that are relevant to the user need. For an efficient Information Retrieval System (IRS) these two sets must be equal as often as possible. The relevance of a document to a query is usually interpreted by most of IR models, vector space [14], probabilistic [12][13][18], inference and belief networks [20][11][17], as a score computed by summing the inner products of term weights in the documents and query representations.
Whatever the used model, the response to a user need is a list of documents ranked according to a relevance value. Many approaches consider term weights as probability of relevance. In such models the incompleteness of information is not considered when representing or evaluating documents given a query. Notions of certainty or possibility are not distinguished in this relevance computing. Yet, the rough nature of document descriptions (a multi set of terms) and of the query description (a list of terms) is hardly compatible with the high precision of relevance values obtained by current methods. The aim of this work is to

propose an IR model based on possibilistic networks. Instead of using a unique relevance value, we propose a possibilistic approach for computing relevance. The relevance of a document to a given query is measured using two values i.e. the necessity and the possibility of relevance. The possibility of relevance is meant to eliminate irrelevant documents (weak plausibility). The necessity of relevance focuses attention on what looks very relevant.
We briefly discuss the use of Bayesian networks in Information Retrieval in section 2. In section 3 we present a general possibilistic approach for IR. We separate reasons for rejecting a document as irrelevant from reasons for selecting it by means of two evaluations: possibility and necessity. This approach is a significant extension of a previous attempt based on possibilistic networks [3]. This extension results from difficulties to find an efficient way of querying the system. It is too restrictive (and demanding) to aggregate query terms by an AND operator when the only information we have is a set of terms. Thus, the idea is to aggregate query terms by conjunction or disjunction operators according to different aggregation methods when no information is given about the logical description of the query. To provide for such a flexibility, a query node is required in the model architecture. We discuss in the latter section the experiments we carried out, showing the importance of weighting schemes we use and by comparing our approach by existing known models on a realistic benchmark.

2 Related works

We discuss in this section the use of Bayesian networks [9][7] in IR, with a view to later comparing it to our model. Bayesian Nets (BNs) [7] provide an efficient tool for storing and reasoning from large probability distributions involving many discrete variables. When probability measures depend on a subjective view, probabilities do not necessarily interpret relative frequencies (related to chance events) but account for degrees of belief (conditional or not). BNs have been used in IR since 1990. The well known IR models using BNs are Inference Networks (INs) and Belief Networks. INs are used in INQUERY system [20] and their efficiency is related to distinct IR approaches and their combination in one model. This system evaluates the belief in a document with respect to a query, and a list of weighted documents is retrieved. Belief Networks [11][17] have been used to "model knowledge derived from past queries and combine it with the vector space model" [11]. The ranking of a document is based on the similarity between document d_j and query Q, computing the probability $P(d_j = 1/Q = 1)$. $Q = 1$ and $d_j = 1$ means respectively Q activated and d_j activated. Recent researchers [4] [5], designed the Bayesian Network Retrieval Model, with a flexible topology that can take into account term relationships as well as document relationships.
The meaning of document and query representations for all these models and relevant document retrieval is identical. For these models a unique degree of relevance is computed and generally weights given to arcs when term nodes are instantiated are based on a combination of $tf - idf$. However, the model we

propose provides a different meaning to document and query representations as well as to the selection of a document given a query. One way to solve our key issue can be given by the use of possibilistic networks.

3 The possibilistic model

One main original idea behind our possibilistic model concerns the relevance interpretation. Instead of using a unique relevance value of a document with respect to a query, we propose a possibilistic approach [6]. A possibility distribution π is a mapping from U to $[0, 1]$. $\pi(u)$ evaluates the plausibility that u is the actual value of some variable to which π is attached. $\pi(u) = 0$ means that u is impossible but $\pi(u) = 1$ only indicates a lack of surprise about u. A proposition A is evaluated by its degree of possibility $\Pi(A) = max_{u \in A}\pi(u)$ and its degree of necessity (or certainty) $N(A) = 1 - \Pi(\overline{A})$ where $\overline{A}$ is the complement of A [6].

3.1 Model architecture

Our approach is based on possibilistic directed networks [1][2], where relations between documents, query and term nodes are quantified by possibility and necessity measures. The proposed network architecture appears on Figure (1). From a qualitative point of view, nodes in the graphical component represent query, index terms and documents and the graph reflects the (in)dependence relations existing between nodes. Document and query nodes have binary domains. A document D_j is invoked or not, taking its values in the domain $\{d_j, \overline{d_j}\}$. The activation of a document node, i.e. $D_j = d_j$ (resp. $\overline{d_j}$) means that a document is relevant or not. A query Q takes its values in the domain $\{q, \overline{q}\}$. As only the positive query instantiation is of interest, we consider $Q = q$ only, and denote it as Q. The domain of an index term node T_i, is $\{t_i, \overline{t_i}\}$. $(T_i = t_i)$ means a term t_i is present in the object (document or query) and thus is *representative* of the object to a certain degree. A *non-representative* term, denoted by $\overline{t_i}$, is a term absent from the object.

Let $\mathcal{T}(D_j)$ (resp. $\mathcal{T}(Q)$) be the set of terms indexed in document D_j (resp. in the query). The query expresses a request for documents containing some terms but excluding other terms. Arcs are directed from document node to index term nodes defining dependence relations existing between index terms and documents. The values taken by index term nodes depend on the document node (parent) instantiation. The query instantiation only gives evidence to propagate through invoked terms, thus arcs are directed from term to query nodes. The terms appearing in the user query form the parent set of Q in the graph. There is an instantiation of the parent set $Par(Q)$ of the query Q that represents the query in its most demanding (conjunctive) form. Let θ^Q be such an instantiated vector. Any instance of the parent set of Q is denoted θ. We show, later in this section, how values are assigned to arcs. For simplicity a query is supposed to contain positive terms only.

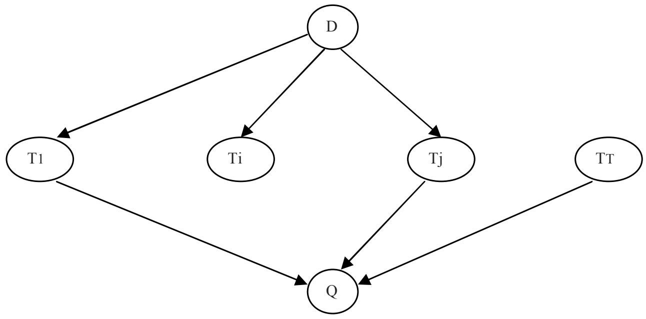

Fig. 1. Model architecture

3.2 Evaluation process

In this model, the propagation process is similar to the probabilistic Bayesian propagation [1][2]. The query evaluation consists in the propagation of new evidence through activated arcs to retrieve relevant documents. Our model should be able to infer propositions like:

- It is plausible to a certain degree that the document is relevant for the user need, denoted by $\Pi(d_j \mid Q)$
- It is almost certain (in possibilistic sense) that the document is relevant to the query, denoted by $N(d_j \mid Q)$

The first kind of proposition is meant to eliminate irrelevant documents (weak plausibility). The second answer focuses attention on what looks very relevant. Under a possibilistic approach, given the query, we are thus interested in retrieving necessarily or at least possibly relevant documents. Thus, the propagation process evaluates the following quantities

$$\Pi(d_j \mid Q) = \frac{\Pi(Q \wedge d_j)}{\Pi(Q)},\ N(d_j \mid Q) = 1 - \Pi(\overline{d_j} \mid Q) \text{ where,} \quad \Pi(\overline{d_j} \mid Q) = \frac{\Pi(Q \wedge \overline{d_j})}{\Pi(Q)}$$

The possibility of Q is $\Pi(Q) = max(\Pi(Q \wedge d_j), \Pi(Q \wedge \overline{d_j}))$ so that $\Pi(d_j \mid Q) = min(1, \frac{\Pi(Q \wedge d_j)}{\Pi(Q \wedge \overline{d_j})})$ [6][1]

We are interested in defining $\Pi(Q \wedge D_j)$. Given the model architecture, it is of the form:

$$\max_{\theta}(\Pi(Q \mid \theta) \cdot \prod_{T_i \in \mathcal{T}(Q) \wedge \mathcal{T}(D_j)} \Pi(\theta_i \mid D_j) \cdot \Pi(D_j) \cdot \prod_{T_k \in \mathcal{T}(Q) \setminus \mathcal{T}(D_j)} \Pi(\theta_k)) \quad (1)$$

for θ being the possible instances of the parent set of Q, θ_i is the instance of T_i in θ. This is computed for $D_j \in \{d_j, \overline{d_j}\}$. Note that terms $T_i \in \mathcal{T}(D_j) \setminus \mathcal{T}(Q)$ are not involved in this computation.

The top retrieved documents are those having a necessary relevance value greater than 0, and the set of possibly relevant documents are retrieved as a second choice.

4 Query aggregation

The possibility of the query given the index terms depend on query interpretation. Several interpretations exist, whereby query terms ($\Pi(Q \mid \theta)$) are defined as expressing *conjunction*, *disjunction*... or, like in Bayesian probabilistic networks, by *sum* and *weighted sum* as proposed for example in the works of Turtle [20]. The basic idea is that for any instantiation θ, the conditional possibility $\Pi(Q \mid \theta)$ is specified by some aggregation function merging elementary possibilistic likelihood functions $\Pi(Q \mid \theta_i)$. Each $\Pi(Q \mid \theta_i)$ is the weight of instance θ_i in view of its conformity with the instantiation of T_i in the query (in θ^Q). We do not consider relations that may exist between terms even if the use of networks would make it possible. Hence, it is difficult (space and time consuming) to store all possible query term configurations or to compute them when the query is submitted to the system. A reasonable organization is to let each query term bear a weight and to compute the weight of joint terms in the query. When the user does not give any information on the aggregation operators to be used, the only available evidence one can use is the importance of each query term in the collection. This evidence is available for single terms that form the query. We give in what follows different manners to aggregate query terms.

4.1 Conjunctive, disjunctive and quantified aggregations

For a Boolean *AND* query, the evaluation process searches documents containing all query terms. Then, $\Pi(Q \mid \theta_i) = 1$ if $\theta_i = \theta_i^Q$, and 0 otherwise. The possibility of the query Q given an instance θ of all its parents, is given by $\Pi(Q \mid \theta)$, where $\Pi(Q \mid \theta) = 1$ if $\forall T_i \in Par(Q)\ \theta_i = \theta_i^Q$ means that the term T_i in θ is instantiated as in the query. Generally this interpretation of the query is too demanding.
For a Boolean *OR* query, the document is already somewhat relevant if there exists a query term in it. The final document relevance should increase with the number of present query terms. The pure disjunctive query is handled by changing $\forall$ into $\exists$ in the conjunctive query. But this interpretation is too weak to discriminate among documents.
Assume a query is considered satisfied by a document if they have at least K common terms. Consider an increasing function, $f(\frac{K(\theta)}{n})$, where $K(\theta)$ is the number of terms in the query instantiated like in a given configuration θ of $Par(Q)$, given that the query contains n terms. It is supposed that $f(0) = 0$ and $f(1) = 1$. f is a fuzzy quantifier [22]. For instance, $f(i/n) = 1$ if $i \geq \frac{K(\theta)}{n}$, and 0 otherwise, requires that at least K terms in the query are in conformity with θ. But more generally f can be a non-Boolean function.
The quantifier approach to computing the possibility of the query Q given an instance θ of all its parents, is given by:

$$\Pi(Q \mid \theta) = f(\frac{K(\theta)}{n}) \qquad (2)$$

4.2 Noisy OR

In general, we may assume that the conditional possibilities $\Pi(Q \mid \theta_i)$ are not Boolean-valued, but depend on suitable evaluations of terms t_i. A possible query term combinations can be "noisy-Or" [9] based. It means that $\Pi(Q \mid \theta)$ is evaluated in terms of conditional possibilities of the form $\Pi(Q \mid t_i \wedge_{k \neq i} \overline{t_k})$ using a probabilistic sum. The primitive terms in a noisy OR are $\Pi(Q \mid t_i \wedge_{k \neq i} \overline{t_k}) = \frac{idf_i}{N} = nidf_i$, denoted $1 - q_i$ for simplicity. Then

$$\begin{aligned} \Pi(Q \mid \theta) &= 0 \text{ if } \nexists \text{ i s.t. } \theta_i = \theta_i^Q \\ &= \frac{1 - \prod_{i:t_i=\theta_i=\theta_i^Q} q_i}{1 - \prod_{T_k \in Par(Q)} q_k} \text{ otherwise} \end{aligned} \quad (3)$$

Only positive terms in the query configuration appear on the numerator. The more query terms present in the document with the same positive instantiation as in the query is, the higher the relevance of the document will be[1].

5 Arc values

In the first part of this section, we define term-document arc values depending on term instantiations. Then we propose a weighting scheme for root nodes. We also weight query-term arcs to aggregate query terms. For the proposed approach we give a weight to prior document possibility, not a uniform one like in inference network model [20] but based on its length.

5.1 Document-term arcs

To evaluate the possibility and necessity of a document relevance we need to express and define relevance represented by arcs in the network. Our approach tries to distinguish between terms which are possibly representative of documents (whose absence rules out a document) and those which are necessarily representative of documents, i.e. terms which suffice to characterize documents.
Postulate 1: A term is all the more possibly representative of a document as it appears frequently in that document;
Postulate 2: A term is all the more necessarily representative of a document as it appears more frequently in that document and it appears fewer times in the whole collection.
According to Postulate 1, $\Pi(t_i/d_j)$ can be estimated from the frequency *tf*:

$$\Pi(t_i/d_j) = nf_{t_{ij}} \quad (4)$$

where $nf_{t_{ij}}$ is normalized term frequency, $nf_{t_{ij}} = \frac{tf_{ij}}{max_{\forall_{t_k \in d_j}}(tf_{kj})}$.
A term weight 0 means that a term is not compatible with the document. If it

[1] We assume Closed World Assumption (CWA): $\Pi(Q \mid t_i) = \Pi(Q \mid t_i \wedge_{k \neq i} \overline{t_k})$

is equal to 1, then the term is possibly representative or relevant to describe the document. Here, "representative" should not be necessarily understood in the general sense, but only as "useful to retrieve this document in the collection". If a term is representative of a document in the general sense, it may not be of much help to retrieve a document. Namely, for a document in a collection devoted to fuzzy sets, the word "fuzzy" is very representative, but it is potentially useless as it does not characterize it among other documents in the same area. Note that the possibility degree is normalized (its maximum is 1). A term not appearing in a document is not compatible with it, and if it appears with a maximal frequency it is considered as a possible candidate to represent it. A discriminant term in a collection is a term which appears (often) in few documents of the whole collection. We assume that a discriminant term is a term which is necessarily representative of a document thus certainly contributes to selecting a document. We define the necessary relevance degree, ϕ_{ij}, of term i to represent a document j as a weight of the form:

$$\phi_{ij} = \mu_1 \left(\frac{N}{n_i} \right) * \mu_2 \left(nf_{t_{ij}} \right) \tag{5}$$

where $*$: product operator and μ_1, μ_2: normalization functions. For instance, μ_1 logarithmic function, μ_2 identity function, and then $\phi_{ij} = \frac{log \frac{N}{n_i}}{log(N)} \cdot ntf_{ij}$ This degree of necessarily relevance shows the necessity for a term to imply a document and thus works to retrieve a document by:

$$N(t_i \to d_j) = \phi_{ij} \tag{6}$$

Since, $\Pi(\overline{d_j}) = 1$ a priori, $\Pi(t_i \mid \overline{d_j}) = \Pi(t_i \wedge \overline{d_j}) = 1 - N(t_i \to d_j) = \phi_{ij}$, while $\Pi(\overline{t_i} \mid \overline{d_j}) = 1$. In table 1, we summarize the conditional possibilities of term instantiations given the document instantiations.

Table 1. Conditional possibility table $\Pi(T_i \mid D_j)$.

	d_j	$\neg d_j$
t_i	$nf_{t_{ij}}$	$1 - \phi ij$
$\neg t_i$	1	1

5.2 Root terms

Weights assigned to terms are mostly the result of a frequentist view because no other information is available. Several works in the literature focus on the definition and on the valuation of the term importance among a collection of documents [12] [18] [8] [10]. Those problems are dealt with, using semantic [21] or a statistical [14][19], [8][23], or probabilistic [10] point of views.

For our approach, when computing the relevance degree to a document given a query, weights must be assigned not only to common terms between the document and the query but also to terms that are present in the query and absent from the document. To be sure to pick up the relevant set of documents, terms must have a discrimination power. The more important is the discrimination power assigned to a term the more efficiently this term helps in the retrieval of documents. This power depends on the distribution of the term among the collection and this distribution is quantified by the density of this term in documents or by the importance of terms across the collection. The less peaked is the density distribution, the less discriminant is the term. We define a new discriminative factor based on entropy, denoted by df_i for a term i in a collection. It improves over the usual idf. The notion of entropy was firstly proposed to evaluate how peaked is a density [16]. The weighting scheme aims to maximize the entropy of the density of the term across the collection.

$$df_i = -\sum_j p_{ij}\ log\, p_{ij}; \qquad (7)$$

$$p_{ij} = \frac{\frac{tf_{ij}}{l_j}}{\sum_k \frac{tf_{ik}}{l_k}}$$

The lower the df_i value is, the more picked the density distribution, and the more interesting the term for retrieving some documents. Thus:

$$\forall t_i \notin \mathcal{T}(D_j),\ \ \Pi(\theta_i) = 1 \text{ if } \theta_i^Q = \overline{t_i}$$
$$= \frac{df_i}{max_{k \in T} df_k} = ndf_i \text{ otherwise} \qquad (8)$$

where T is the set of terms in the collection.
The more discrimination power the term has, the less documents d_j not containing it are relevant to the query and the lower is $\Pi(Q \wedge d_j)$. It is clear that two documents with same $idf's$ may have different df values. idf is less discriminant than df.

5.3 Prior possibility of documents

In absence of information, the *a priori* possibility on a document node is uniform ($\Pi(d_j) = \Pi(\overline{d_j}) = 1$). Actually, we can obtain information on a document given the importance of its terms, its length etc. This knowledge can be given for instance, by the user, the user profile etc. Hence, for example, if we are interested in retrieving long documents, we define the prior possibility of activating a document $D_j = d_j$, $\Pi(d_j) = \frac{l_j}{max_{k=1,..,N} l_k} = nl_{d_j}$ where l_j is the length in frequency of document d_j; $l_j = \sum_i tf_{ij}$. The shorter the document is, the less possibly relevant it is. Besides, $\Pi(\overline{d_j}) = 1$.

6 Experiments and results

The experiments were undertaken on the dataset *Le Monde*. The aim of these experiments is to evaluate the reliability of the proposed approach based on different weighting schemes specifically one considering the term discriminative power based on entropy *ndf* and the second one the known *nidf* and its evaluation process, i.e. its evaluation of documents given a query, based on two measures of relevance. The results obtained are discussed below and then compared to OKAPI's weighting scheme.

6.1 The dataset collection

Experiments are carried out on a sub-collection of CLEF, namely LeMonde 1994, a French newspapers collection (44013 documents and 34 queries, 154 MB). For each query, 1000 top documents are retrieved. We give the results evaluation by means of $P5, P10, ...$ i.e. the precision at point 5, which is the ratio of relevant documents among the 5 top retrieved documents, among the 10 ones, etc.

6.2 Parameters

Our model is based on two measures that evaluate two kinds of relevance i.e. the necessary and the possible one. The trust in necessarily relevant documents is greater than in possibly relevant documents. If less than 1000 necessary documents are retrieved we complete to 1000 by adding possibly relevant documents.
As shown in sections above and by formula 1, different pieces of information

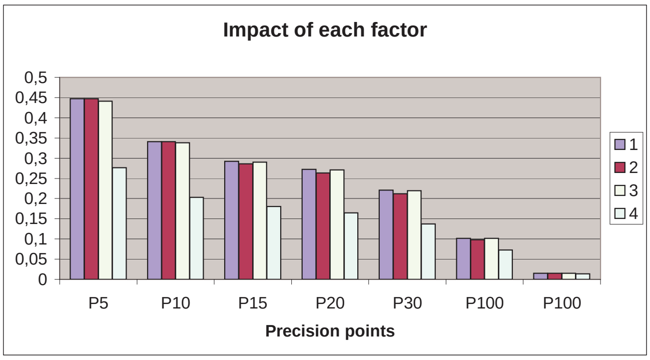

Fig. 2. Weights impact

are used for the document evaluation given the query: the distribution of terms inside a collection, the *a priori* possibility of documents, the importance of query term absent from documents (*ndf* factor). The tuning of parameters is described in table 2. *Yes* indicates the parameter is used and *No* that it is not used in the computations. Figure 2 shows the impact on precision points of each piece of information. As example, in case 1, the normalized length(nl_d), the normalized

Table 2. Parameters tuning

	nl_d	ndf'	Noisy Or	ndf
1	Yes	Yes	Yes	No
2	Yes	No	Yes	Yes
3	Yes	Yes	No	No
4	No	Yes	Yes	No

entropy ($ndf' = \frac{ndf}{max_{t \in \mathcal{T}(\mathcal{Q})} ndf_t}$) and the Noisy Or are used for the document evaluation given the query. The ndf' as shown in section 5.2 is normalized given all terms of the collection and a second time given the query terms. When not *ndf* nor ndf' factors are considered inside the computations the results decrease strongly (about 90% less for average precision)[2]. It is the case especially for query terms having a high potential discrimination power between documents, i.e., terms which have high density in few documents of the collection. In a such case, there are no necessary documents for an important number of queries. This *ndf* factor strongly decreases the relevance of documents not containing "interesting" terms. In the case of removing the documents length, from the propagation process, the number of necessary relevant documents increases. This is because short documents relevance grows up as they contain not interesting terms: documents with higher *ndf* (for root terms) than *nidf* (for present terms). When the Noisy or is not considered, weights affeted to query terms present in documents equal 1.
The average precision is higher when *nidf* is kept out computations than when *ndf* is removed from the propagation computations. In our model, the *df* factor is used once to decrease the relevance of documents not containing "interesting" terms, whereas *nidf* factor is used for terms present in the document under concern. Both factors (*nidf* and *ndf*) try to find the extent to which a term is specific in a given collection, the *nidf* in terms of presence/absence (of a term in a document) whereas *ndf* in terms of density distribution.

6.3 Comparison

Figure 3 shows the precision points comparison between our approach (*Pi*-nets) and the probabilistic approach (*OKAPI*).

The comparisons is between the BM25 weighting scheme (OKAPI) [12][13] and our approach. We can note from figure above (figure 3) that precision points are better for our approach for any points of precision. It improves average precision by 7.99% compared to OKAPI system.

[2] This result does not appear in the figure 2.

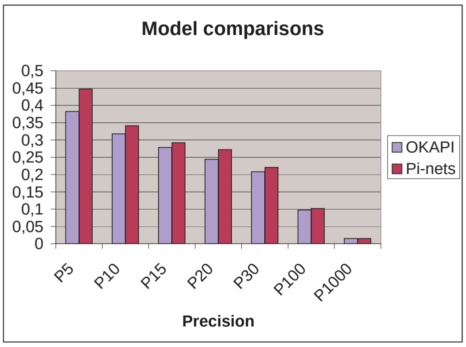

Fig. 3. Models comparison

7 Conclusion

This paper presents a new IR approach based on possibility theory and a new term discrimination index based on entropy. In a general way, the possibility measure is convenient to filter out documents (or index terms from the set of representative terms of documents) whereas necessity captures document relevance (or index terms representativeness). The originality of the proposed approach is due to the use of information about the distribution of terms across the collection (example: use of df), and a new way of indexing documents by separating different kinds of information. The first experiments carried out on LeMonde are promising. Other experiments on the benchmark $WT10G$ of TREC are also promising as we seem to obtain in most cases better average precision than known models.

Aknowledgments

Thank to Luis de Campos for helpful discussions on the original possibilistic model

References

1. S. Benferhat, D. Dubois, L.Garcia, H. Prade: Possibilistic logic bases and possibilistic graphs. In Proc. of the 15th Conference on Uncertainty in Artificial Intelligence, (1999) 57-64.
2. C. Borgelt, J. Gebhardt and R. Kruse: Possibilistic graphical models. Computational Intelligence in Data Mining, CISM Courses and Lectures 408, Springer, Wien, (2000) 51-68.
3. A.H. Brini and M. Boughanem and D. Dubois: Towards a possibilistic approach for Information Retrieval. Proc. EUROFUSE Data and Knowledge Engineering, Warsaw, (2004) 92-102

4. L.M. Campos, J.M. Fernandez-Luna, J.F. Huete. The BNR Model: Foundations and Performance of Bayesian Network-based Retrieval Model. JASIST, (2003) 54(4),302-313.
5. F. Crestani, L.M. de Campos, J.M. Fernandez-Luna, J.F. Huete: A Multi-layered Bayesian Network Model for Structured Document Retrieval, Proc. ECSQARU LNAI 2711, Springer, (2003) 74-86.
6. D. Dubois and H. Prade: Possibility theory. Plenum, (1988)
7. F.V. Jensen. Bayesian Networks and Decision Graphs, Springer, 2000.
8. H.P. Luhn: The automatic creation of literature abstracts. IBM Journal of Research and Development, (1958), 2, 159-165.
9. J. Pearl: Probabilistic Reasoning in Intelligent Systems. Morgan Kaufmann San Mateo, Ca, (1988)
10. M.E. Maron: Automatic indexing: an experimental enquiry. Journal of the ACM, (1961), 8, 404-417
11. B. Ribeiro-Neto, I. Silva, R. Muntz: A Belief Network Model for IR. In Proc. Of the 19th ACM SIGIR Conf. on Research and Development in Information Retrieval, (1996), 253-260, Zurich, Switzerland.
12. S.E. Robertson and S. Walker: Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval, Proc. of the 17th Ann. Inter. ACM SIGIR Conf. on Research and Dev. in Information Retrieval, Springer-Verlag, (1994) 232-241
13. S.E Robertson, S. Walker, S. Jones, M.M Hancock-Beaulieu and M. Gatford: Okapi at TREC-3. Proc. of the 3rd Text REtrieval Conference (TREC-3), NIST Special Publication (1995) 109-126
14. G. Salton: The Smart retrieval system-experiments. In Automatic Document Processing, Prentice Hall Inc, (1971)
15. G. Salton and C. Buckley: Term-weighting approaches in automatic text retrieval. IPM, (1988) VOL 24, 513-523
16. C. E. Shannon: The mathematical theory of communication. Bell System Technical Journal, (1948) Vol. 27, 379-423 and 623- 656.
17. Ilmerio Silva, Berthier Ribeiro-Neto, Pavel Calado, Edleno Moura, and Nivio Ziviani: Link-Based and Content-Based Evidential Information in a Belief Network Model. In ACM SIGIR 23rd Int. Conference on Information Retrieval, Athens, Greece, (2000), 96-103.
18. K. Sparck Jones, S. Walker and S.E. Robertson: A probabilistic model of information retrieval: development and comparative experiments, Parts 1 & 2. IPM, (2000) VOL 36 779-808,809-840
19. K. Sparck Jones: A statistical interpretation of term specificity and its application in retrieval. Journal of Doc, (1972) VOL 28 111-121
20. H.R. Turtle and W.B. Croft: Inference networks for document retrieval. In Proc. 13th Int.Conf. on Research and Development in Information Retrieval. (1990) 1-24
21. K. Van Rijsbergen: A Theoretical basis for the use of co-occurrence data in information retrieval. In Jour. of Doc. (1977) 33 106-119.
22. R. R. Yager and H. Legind Larsen: Retrieving information by fuzzification of queries. Int. Jour. of Intelligent Inf. Systems, (1993) Vol. 2(4).
23. H.P. Zipf: Human behaviour and the principle of least effort. Addison-Wesley, Cambridge, Massachusetts (1949).