Using Plausible Inferences and Dempster-Shafer Theory of Evidence for Filtering

Farhad Orumchian, Babak Nadjar Araabi
Control and Intelligent Processing Center of Excelence
Department of Electrical and Computer Engineering, University of Tehran, Iran
e-mail: foroumchian@acm.org , araabi@ut.ac.ir

Elham Ashoori
Control and Intelligent Processing Center of Excelence
Department of Electrical and Computer Engineering, university of Tehran, Iran
e-mail: ashoori@iman.ut.ac.ir

Abstract*: This paper describes a set of experiments investigating the use of Dempster-Shafer (DS) theory of evidence in combining the evidences provided by plausible inferences in Information Retrieval. PLIR is an experimental information retrieval system based on the theory of plausible reasoning[1] which tries to simulate a reference librarians reasoning while trying to retrieve documents in response to queries. Often a document can be retrieved by several plausible inferences. Each inference constitutes of an uncertain evidence of relevance. In these experiments, DS theory is used in several ways to combine these evidences in PLIR system. The application of DS theory and weighting inferences based on their overall usefulness in retrieval has improved the quality of ranking of PLIR system. We use this conclusion in our future research in the application of plausible inferences and DS theory in adaptive information filtering.*

Keywords: Plausible Reasoning, Dempster-Shafer theory of evidence, Information Filtering, Information Retrieval, High Precision.

1. Introduction

Our experiments focus on the problem of combining the evidences regarding relevance of the documents provided by means of Plausible Reasoning from the content of documents and knowledge of the corpus. In our system, every query consists of one or more terms. Each query term can be a single word, phrase or a logical term. A logical term is a logical representation of a relation between two single words or phrases. For example, the sentence fragment "query language for a relational database" describes a relationship between two phrases "query language" and "relational database". In PLIR, this relationship is captured in a logical term as "query_language (relational_database)". PLIR utilizes different inferences of plausible reasoning to retrieve a number of documents along with a degree of confidence in their relevance. Each document can be retrieved with multiple inferences. Many documents could be retrieved for each term in the query. Our problem is how to combine the results of different inferences for each query term and how to combine the results of different terms of the query.

We need two kinds of aggregation functions at these two levels. Combining evidences of relevance from several inferences for a single query term is a matter of finding the right theory and parameter settings. But the problem of combining relevance evidence from several query terms is not as simple. That is because many external parameters such as user preferences could influence the combination of the evidences provided by inferences. The different levels of combination are shown on Fig 1.

In the PLIR system, if no document can be inferred from a query term, it only means that the system is unaware of any evidence about relevance of any document to this term. In PLIR system, There are two approaches in combining these evidences. The global approach is to combine all the evidences provided by all the inferences on all the terms of the query together in one step. The second approach is a local approach. First we combine the relevance evidences from each inference on a term to get a degree of relevance for each document inferred from that term. Then in second step, we combine all these degrees of relevance of the documents, provided by different terms, to estimate the overall degrees of relevance of the documents to the whole query. In our experiments we only considered the latter approach.

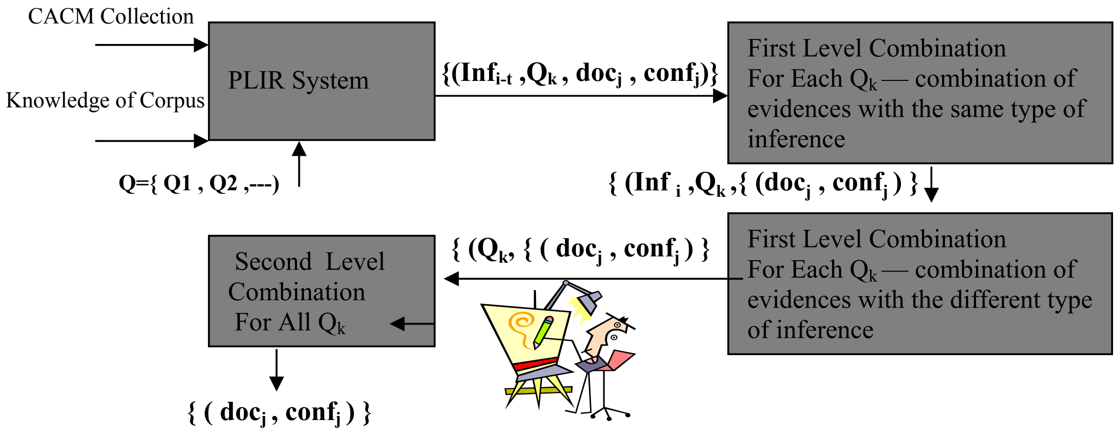

Figure 1. Problem Statement

For the first level combination, we assume that the evidences are supportive, so we must use an aggregation function that increases the confidence in relevance; in another words, ***confidence (A ^ A) <> confidence (A)***. The more evidence the higher confidence in relevance. However, care must be taken in the choice of the combination function. The function should be sensitive to the different confidence values assigned to a document from different query terms. For example, assume a query has two terms q1, q2 and there are two plausible inferences. Every inference returns a confidence value for each document corresponding to each term of the query. Table 1 describes three different cases. In each case first level results are combined with both addition and Dempster's rule of combination. At the second level, it is assumed that the user has no preferences for either of the query terms therefore, (s)he would be satisfied with the mean of the confidence values of the query terms.

Comparing the results of second layer for the three cases, when Dempster's rule of combination is used, reveals that results are sensitive to the permutation of the confidence values for all query term. As can be seen in the table, user would be most satisfied with case#3. On the other hand, we can see that when using simple addition for the first level, the results of this level are larger than every individual confidences, but these results are not good enough to make the distinction among three cases in the second level.

Based on this observation, a number of experiments were carried out, investigating the effectiveness of the application of Dempster's rule of combination at the first level.Section 2 introduces briefly the DS theory of evidence in the context of our problem. Section 3 discusses the method of scoring of documents in PLIR system. Section 4 presents the experiments.carried out on CACM collection. Section 5 describes an analysis of the results. Section 6 introduces Information Filtering by Plausible Reasoning. Section 7 is the conclusion and future work.

2. Dempstr-Shafer Theory of Evidence

In this section we describe the main concepts of DS theory, within the context of our problem. DS theory of uncertainty was first introduced by the statistician Arthur Dempster [2] and extended by Glenn Shafer[9]. Theory may be considered as a generalization of the probability theory, with two differences: First, the possibility of explicit representation of ignorance; Second, the combination of evidences. With Dempster's rule of combination, the evidences from multiple sources can be combined with each other. Examples of applications of this theory in Information Retrieval can be found in [3,5,6,7,8,12]. In the context of this problem, uncertainty refers to the following three cases: First, the existence of multiple evidences for the relevance of a document to a query term. Second the amount of unspecified evidences for the relevance of that document to the same query term. Third, the misleading evidences that incorrectly identify the document as relevant to that query term. In our problem the evidences are compatible with each other and have no conflict.

2.1 Frame of Discernment

The DS framework is based on the view whereby propositions are assumed as subsets of a given set of hypotheses [9]. In PLIR system, the documents are retrieved through plausible inferences on query terms. The set of all the possible answers to the query that are documents is Ω. The set is called a frame of discernment. In our problem, the frame of discernment is taken to be the set of all documents in the CACM collection.

Table 1 . An example of combining inferences with DS theory and addition in first level and arithmetic mean at second level

Inference	Query Term	Confidence	First Level DS Rule	Second Level Arithmetic Mean	First Level Add	Second Level Arithmetic Mean
Case # 1						
1	Q1	0.3	0.37	0.317	0.4	0.34
1	Q2	0.2	0.264		0.28	
2	Q1	0.1	0.37		0.4	
2	Q2	0.08	0.264		0.28	
Case # 2						
1	Q1	0.3	0.44	0.306	0.5	0.34
1	Q2	0.1	0.172		0.18	
2	Q1	0.2	0.44		0.5	
2	Q2	0.08	0.172		0.18	
Case # 3						
1	Q1	0.3	0.356	0.318	0.38	0.34
1	Q2	0.2	0.28		0.3	
2	Q1	0.08	0.356		0.38	
2	Q2	0.1	0.28		0.3	

Table 2. Different Experiments on Combining Evidences

Exp#	$m_i(doc_j)$	$m_i(T)$	$m_i(\phi)$	m (doc_j)
Exp#1	c_i	$1-c_i$	0*	$m1 \otimes m2 \otimes m3 \otimes ...m5$
Exp#2	$w_i * c_i$	$w_i * (1-c_i)$	< >0*	$m1 \otimes m2 \otimes m3 \otimes ...m5$
Exp#3	$w_i * c_i$	1- $w_i * c_i$	0*	$m1 \otimes m2 \otimes m3 \otimes ...m5$
Exp#4	$w_i * c_i$	$w_i * (1 - w_i * c_i)$	< >0*	$m1 \otimes m2 \otimes m3 \otimes ...m5$
Exp#5	$w_i * c_i$			$m1 + m + \otimes m3 + ...m5$
Exp#6	$w_i * c_i$	$w_i * (1-c_i)$	< >0**	$m1 \otimes m2 \otimes m3 \otimes ...m5$

2.2 Basic Probability Assignment

Every inference returns a confidence value for each document inferred by each term of the query. These confidence values are modeled by a density function $m: 2^{\Omega}$ → [0,1] called a basic probability assignment (bpa).

$$m(\phi) = 0, \sum_{A \subset \Omega} m(A) = 1$$

m(A) represents the belief exactly committed to A, that is the exact evidence that the document is relevant to a query term. If there is positive evidence for relevance of a document to a query term, then m(A) > 0, and A is called a focal element. The focal elements and the associated bpa define a body of evidence. In this problem, we assume that focal elements are singleton sets. If m(A) = 0 then there is no confidence about non relevance of a document to any particular query term. Each body of evidence is composed of the confidence on relevance of a document to each query term as estimated by inferences of plausible reasoning. Then

$$m(\phi)=0, m(\{doc_j\})+m(T)=1$$

m(T) is referred to evidence that can not be assigned yet (uncommitted belief as described in [3]). The m(T) represents the uncertainty (overall ignorance, lack of knowledge) associated to the entire set of documents about being relevant to a query term.

2.3 Belief Function

Given a bpa m, belief function is defined as the total belief provided by the body of evidence for relevance of a document to a query term. Because the focal elements are singleton, then the belief function equates to the mass function.

2.4 Dempster's rule of combination

Dempster's rule of combination aggregates two independent bodies of evidence defined within the same frame of discernment into one body of evidence. Let m1 and m2 be the bpas associated to a document for relevance to a term. Because the focal elements are singleton, the combination function becomes simpler than Dempster's rule of combination. Only the evidences with $m_i(doc_j) > 0$ combine with each other.

$$m(doc_j) = m1 \otimes m2(doc_j) = (m1(doc_j) * m2(doc_j) + m1(doc_j) * m2(T) + m2(doc_j) * m1(T)) / K$$

$$m(T) = m1(T) * m2(T) / K$$

K is a normalization factor to support that m is bpa

$$K = m1(doc_j) * m2(doc_j) + m1(doc_j) * m2(T) + m2(doc_j) * m1(T) + m1(T) * m2(T) + m1(\phi) * m2(\phi)$$

Where

$$m(doc_j) + m(T) + m(\phi) = 1$$

3. Scoring of documents

DS provides three functions for scoring of documents: mass, belief and plausibility functions. In this work, For the first level, we compute a mass function to combine the evidences for one query term. In the second level, the evidence of each query part for different document must be combined to compute the final result. In second level the evidences that we want to combine are defined on different frames of discernment. In this level, an adaptive function can be defined to adapt the combination based on the user interests. In these experiments, we assumed that user would be satisfied with the documents with the high degree of relevance to all the query terms. In other words, no preferences are given to any of the query terms, therefore an averaging function would satisfy user's needs.

4. Experiments

The PLIR system was tested on CACM collection and it was demonstrated that it is a high precision retrieval model [4]. This model uses plausible inferences to infer the relevance of documents to query terms. Then it combines the different confidence values estimated by different inferences and ranks the documents based on their overall confidence values. In the mentioned experiments, PLIR used two different approaches for combining these evidences of relevance. The pessimistic approach took the maximum of confidence values of all the inferences pointing to a document as relevant. The optimistic approach used the following formula: $C = C1 + C2 - C1 * C2$

Where C1 and C2 are confidences estimated through two different inferences. Both pessimistic and optimistic approaches are global. That means they compute overall confidence in relevance among all the inferences. Both above approaches demonstrated the same level of performance in terms of precision. A problem that was noticed in the result of both of those methods was that some times they did not distinguish among the documents. They assigned the same confidence value to several documents.

For the current experiments we had two goals in mind: first to improve the ranking by moving relevant documents to higher ranks as much as possible. Second, to distinguish among those documents that have been assigned the same confidence values in the previous experiments.

All of our current experiments are on the methods of combining evidences provided by plausible inferences of PLIR with a local approach. Therefore we used the confidence values produced by PLIR and tried several different approaches in combining them. In current experiments we refer to the previous experimental results as exp#0.

In these experiments, only five types of inferences of PLIR system were used. These are: Referent Based Transforms, Argument Basesd Transforms, Mutual Dependency, Terms and Finding Referent. Different experiments are described in Table2.

First, we define C_i , W_i (weight of inference_i) and $m_i(doc_j)$ (mass of evidences of relevance for doc_j from inference_i)

C_i = confidence on relevance of each document for each term of the query returned by inference_i

W_i = number of relevant documents retrieved by inference_i / total documents retrieved by inference_i

Then

$$W_i = \frac{W_i}{\sum_{i=1}^{5} W_i}$$

$$m(doc_i) = m1 \otimes m2 \otimes m3 \otimes \ldots m5(doc_i)$$

$m1 \otimes m2(doc_i)$ is defined in section 2.

The method of computing uncommited belief:
The null set is assumed empty or not; In Table2, if the $m_i(\phi)$ marked with *, results are normalized; If $m_i(\phi)$ marked with **, the results are not normalized with defined K, But only divided by a smaller value that results in computing K with considering $m_i(\phi) = 0$.

Experiments differ in below 4 aspects:

1- The method of computing confidence on one inference that could be simple or weighted.
2- The combination of evidences of relevance from different inferences for each query term that could be Dempster's rule of combination or addition.
3- The method of computing unassigned certainty
4- The combination of evidences of relevance from different query terms.

In Exp#1 there is no preference among inferences; therefore the confidences are not weighted. The uncommitted belief is one minus the total belief assigned. The differences between this experiment and pessimistic approach in [4] are only the difference between global and local approach but the combination rule is the same. In Exp#2 to Exp#6, the confidence coming from better inferences (inferences correctly identifying relevant documents) are preferred over the confidence coming from other inference. In Exp#2, the weights of inferences are taken into consideration to decrease the uncommitted belief. The mass of null set is not equal to zero, but the end, the results are normalized. In Exp#3, the weights of inferences are taken into consideration to increase the uncommitted belief. The mass of null set is equal to zero. In Exp#4, everything is the same as Exp#2, but the uncommitted belief is more than Exp#2. In Exp#5, the uncommitted belief is not taken into account. For the second level of combination, an addition operator is used to combine the weighted confidences. Exp#6 is similar to Exp#2 but the results of mass functions are not normalized. In this experiment, we have considered the open world assumption [10] to model the capability of the reference librarian in this collection. The open world assumption reflects the idea that Ω might not contain the actual world [11]. So, Instead of normalization, we only do not consider $m_i(\phi)$ in computing K and divide the mass with a value less than the value that used for normalization in other experiments.

5. Results

Exp#0 assigns the same confidence value to a group of documents and other experiments are based on results from Exp#0. For comparing the performances of the experiments with each other and to see how different methods break the ties, a modified version of precision is used. The definition of this precision is as below:

$$RP_i = r_i * p_i / r$$

Where:
r_i : number of retrieved documents at ith rank
p_i : precision at ith rank
r : total number of retrieved documents
This measure prefers experiments that have more relevant documents in higher ranks and ranks have lesser documents assigned to them.

Table 3 contains the query by query results of comparing all the experiments. Figure 2 depicts the same results graphically comparing with Exp#0.

In all experiments the ranking has improved over original experiments. Each one of experiments produced more ranks with lesser number of documents in each rank. In these experiments Exp#6 has produced better results. Since the test data was small and a good number of documents were only retrieved by a single inference, therefore we do not feel comfortable to generalize the results. However, We believe this result may be reproducible in larger experiments. Therefore we are planning to repeat the same tests on *data* collection *of TREC-9 Filtering Track*.

Table 3. The query-by-query results comparing all the experiments

query#	exp# 0	exp# 1	exp# 2	exp# 3	exp# 4	exp# 5	exp# 6
4	0 .53	0 .53	0 .4062	0 .5062	0 .5062	0 .5062	0 .5062
6	0 .75	0 .75	0 .25	0 .75	0 .25	0 .75	0 .75
7	0 .7491	0 .78	0 .7284	0 .7869	0 .7284	0 .7869	0 .7869
8	0 .5027	0 .19	0 .5545	0 .1899	0 .5545	0 .1899	0 .1899
9	0 .75	0 .81	0 .7223	0 .7958	0 .7223	0 .7958	0 .7958
11	0 .4337	0 .47	0 .3247	0 .4629	0 .3375	0 .4629	0 .4845
13	0 .668	0 .67	0 .6816	0 .6816	0 .6816	0 .6816	0 .6816
15	0 .3111	0 .35	0 .2926	0 .3624	0 .2926	0 .3624	0 .5106
16	0 .172	0 .23	0 .0967	0 .2378	0 .1289	0 .2378	0 .2378
17	0 .5131	0 .60	0 .4789	0 .6006	0 .4945	0 .6006	0 .6006
19	0 .75	0 .75	0 .6611	0 .7917	0 .6611	0 .7917	0 .7917
21	0 .1034	0 .10	0 .1166	0 .0953	0 .1166	0 .0953	0 .0953
22	0 .25	0 .54	0 .2083	0 .5833	0 .2083	0 .5833	0 .5833
25	0 .5064	0 .53	0 .4688	0 .5336	0 .4672	0 .5414	0 .5558
29	0 .9051	0 .91	0 .9139	0 .9078	0 .9139	0 .9078	0 .7051
30	0 .4375	0 .44	0 .4509	0 .4509	0 .4509	0 .4509	0 .4509
32	0 .5556	0 .56	0 .1111	0 .5556	0 .1111	0 .5556	0 .5556
40	0 .3503	0 .40	0 .2733	0 .4133	0 .2733	0 .4133	0 .3883
60	0 .75	0 .75	0 .7833	0 .7833	0 .7833	0 .7833	0 .7833
61	0 .2074	0 .23	0 .2074	0 .2351	0 .2074	0 .2351	0 .3007
63	0 .75	0 .75	0 .6611	0 .7917	0 .6611	0 .7917	0 .7917
comparing_measure	0 .4577	0 .47	0 .4254	0 .4697	0 .428722	0 .4698	0 .4703

6. Information Filtering By Plausible Reasoning

following the definition of retrieval process in [4] , filtering process is defined as expanding a query description by applying a set of inference rules continuously on the description of the query and inferring other related concepts, logical terms and statement until locating a document(s) which are described partially by these concepts or logical terms or statements, considering the user's preferences and feedback. One kind of modified inference rules of Plausible Reasoning that can be used in information filtering, is the "document filtering using a GEN-based argument transform" shown here.

REF (d (a)) = { ? }	
1- d (a') = r	γ 1
2 - a GEN a' in CX (a , D (a))	δ 1 , A1
3 - D (a) <---------------→ d (a)	α , γ 2
4 – d (a) = r	γ 3 = F1 (γ 1 , δ 1 , A1 , α , γ 2) from 1,2,3 by Gen-based Argument transform ,B1
5 – REF (r) = {doc#x}	δ 2 , A2
REF (d (a)) = {doc#x}	γ =F2 (γ 3 , δ 2 , A2 , B1)

B1 can be computed individually for each user. It will represent the point of view of a particular user over time about different type of argument "a" retrieved by the system. Details on the description of above parameters can be found in [4].

In information filtering we try to investigate the dominant inference rules and the weight of each inference type of plausible reasoning for individual user. Special methods are suggested for evaluating the value of certainty parameters for combining these values in PLIR system. In this work, we apply Dempster's rule of combination to combine the final value of certainty parameter for all inference rules for each query term.

7. Conclusion

A series of experiments were conducted to improve the quality of ranking of the PLIR system. The PLIR system is an experimental high precision retrieval system based on the theory of the plausible reasoning of Collins and Michalsky. Since PLIR retrieves documents through plausible inferences, these inferences could be considered as sources of evidence of relevance of a document to a query or query term. For these experiments the application of Dempster-Shaffer theory of evidence was considered for combining the evidence gathered through plausible inferences. Several experiments were conducted with different assumptions and settings.

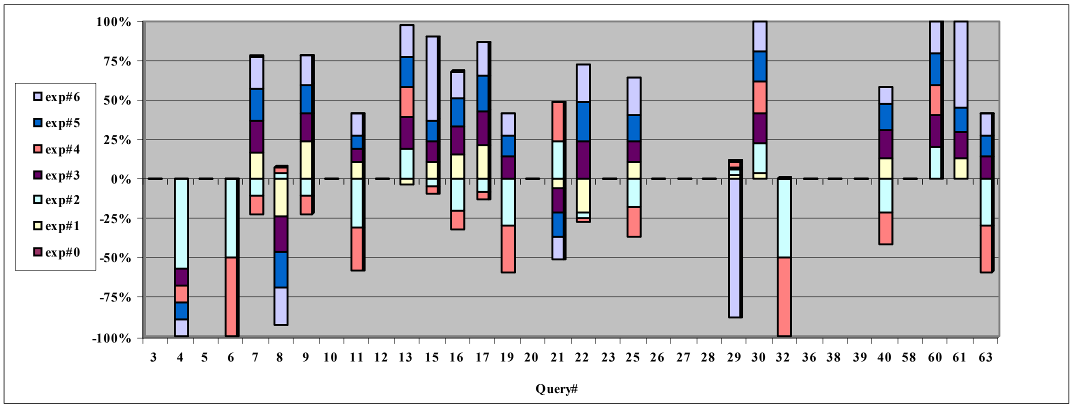

Figure 1. The difference of comparing measure with Exp#0

In general it can be considered that the application of DS theory in combining the evidences produced by plausible inferences produces more qualitative ranking. It seems, this approach pushes the non-relevant documents to lower ranks. Therefore with a good method of computing a query-based threshold, one could eliminate many of non-relevant documents. Since PLIR is a high precision retrieval system that retrieves few documents, it seems reasonable to believe the combination of PLIR with DS theory and a good query-based threshold could be useful for information filtering. Since PLIR allows adaptation to user needs therefore threshold can be learned from user relevance judgments. Then this approach could be more useful for adaptive information filtering.

In these experiments we only touched the misleading inferences problem. Misleading inferences generate misleading evidences that can be taken into account when all evidences of relevance are combined. We also want to experiment with the misleading inferences by using user's relevance feedback and considering their weight in the evidence combination formula.

Another interesting idea that we just started to play with is: the open world assumption. In future we plan to use both ideas in combining evidences and computing the confidence of relevance. We want to concentrate more on smets's open world assumption in future experiments.

We also consider the DS combination function to combine the evidences that affect relevance feedback to update the query instead of utilizing the Rochioo formula.

References

[1] Allen Collins, and R. Michalski. "The logic of plausible Reasoning A core theory", cognitive science, vol. 13, pp.1-49

[2] A. Dempster, "Upper and lower probabilities induced by a multivalued mapping," Ann. Math. Stat., vol. 38, no. 2, pp. 325–339, 1967

[3] M. Lalmas, and M. Ekaterini, "A Dempster-Shafer indexing for focussed retrieval of a hierarchically structured document space: Implememtation and experiments on a web museum collection," 6th RIAO Conference, Content-Based Multimedia Information Access, Paris, France, April, 2000.

[4] F. Oroumchian, and R.N. Oddy, "An Application Of Plausible Reasoning To Information Retrieval," proc. of the 19th ACM SIGIR Conference on Research and Development in Information Retrieval, Zurich, Switzerland, pp. 244-252, August, 1996

[5] I. Ruthven, M. Lalmas, and K.V. Rijsbergen, "Combining and selecting characteristics of information use," Journal of the American Society of Information Science and Technology, 53(5), pp. 378-396, 2002.

[6] I. Ruthven, and M. Lalmas, "Using Dempster-Shafer's Theory of Evidence to combine aspects of information use," Journal of Intelligent Information Systems, 2001

[7] I. Ruthven, and M. Lalmas, "Representing and Retrieving Structured Documents using the Dempster-Shafer Theory of Evidence: Modelling and Evaluation," Journal of Documentation, 54(5), pp. 529-565, December 1998.

[8] I. Ruthven, and M. Lalmas, "Experimenting on Dempster-Shafer's theory of evidence in information retrieval," Technical Technical report, University of Glasgow, April 1998.

[9] G.A. Shafer, Mathrematical Theory of Evidence, Princeton University Press, 1976.

[10] Ph. Smets, The nature of the unnormalized beliefs model, In D. Dubois, M. P. Wellman, B. D'Ambrosio, and Ph. Smets, editors, Uncertainty in Artificial Intelligence 92, Morgan Kaufman, pp. 292–297, San Mateo, CA, 1992.

[11] Ph. Smets, "Belief Functions and The Transferable Belief Model," page at the web site of the Imprecise Probabilities Project: http://ippserv.rug.ac.be. last visited : 31/10/2002

[12] M. Theophylactou, and M. Lalmas, "A Dempster Shafer Model for Document using Noun Phrases," Proc. of the 20th Colloquium of British Computer Society's Information Retrieval Specialist Group. Electronic Workshops in Computing, 1998.