2012 International Conference on Information and Knowledge Management (ICIKM 2012)
IPCSIT vol.45 (2012) © (2012) IACSIT Press, Singapore

N-Order Weight Propagation on Collections with Related Documents

Fadi Yamout

Lebanese International University
Beirut, Lebanon
fadi.yamout@liu.edu.lb

Abstract. N-Order Weight Propagation (NWP) is an enhanced version of the Relevance Feedback (RF) Weight Propagation (WP) technique which moves more relevant documents higher up the ranking. The technique propagates positive and negative weights from documents judged relevant and not relevant respectively, to other documents. Further improvement in precision in WP is mostly attributed to the NWP technique where the documents that received the highest weights, and called top-set, propagate themselves to documents close by in vector similarity based on threshold Epsilon ε. NWP and WP were tested on Web collections with documents that cover various subjects. However, other non-Web collections exist and contain documents related to the same subject. These collections are usually used by regular users, researchers or students searching for information in research centers or libraries related to a specific subject such as Aerodynamic, Medical, and Information Science…etc... . Many of the RF techniques perform well in Web collections and fail in these collections with related documents. In this paper NWP and WP techniques are tested using Text REtrieval Conference (TREC)[1] small test collections of articles from aerodynamics journals (Cranfield), articles from medical journals (Medline), documents on information Science (CISI), abstracts from Library and Information Science journals (LISA) and document titles on electronic engineering (NPL). Various models are used in the experiments such as the vector model using the tf.idf weighting scheme, and the probabilistic model using the DB2 model in the context of the Divergence from Randomness (DFR) framework. In this paper the constant Epsilon ε will be determined for the collections for the various models.

Keywords: Information Retrieval, Relevance Feedback, Vector Model, Probabilistic Model

1. Introduction

One of the most frequent activities by people is becoming searching on the Internet for electronic documents. Search engines are the main tools for this search. During the 1970's databases were computerized and Information Retrieval became more popular since plenty of text was stored. With relevance feedback technique, the user is able to mark which of the retrieved documents were and were not relevant [1]. The reformulated query can be resubmitted to the system, and the search process is repeated. Using negative as well as positive relevance feedback, index terms found in documents judged not relevant are removed from the query.

In this paper we will test a recently developed relevance feedback technique called N-Order Weight Propagation (NWP) [2] on test collections of homogenous nature namely with documents related to the same subject. NWP has shown improvement over the baseline when tested on Web collections with documents of diverse subjects. In addition we will determine the best value for ε used in NWP to increases or reduces the number of documents to propagate.

[1] Information can be found at http://trec.nist.gov/

The structure of the remainder of this paper is as follows: in Section 2 we review the RF techniques and in section 3 the NWP technique. In Sections 4 and 5 we describe the experimental design and results of experiments, and finally conclude in section 6.

2. Relevance Feedback Techniques

The earliest work on relevance feedback by Rocchio [3] was founded on the vector space model which is mainly based on the tf.idf weighting technique [4]. In a vector space model [4], the semantics of the document are derived from the document's component terms. A document is represented as a vector of terms $d_j = (t_{1j}, t_{2j} \dots t_{mj})$ where m represents the number of unique terms in the collection and t_{ij} denotes the presence or absence of term i in document j. Other work on relevance feedback was based on Information-theoretic query expansion [5][6][7] where the topmost documents to be assessed are called "Elite_set T of documents" and the most informative terms are selected by using the information-theoretic Kullback-Leibler divergence function (KL). Information-theoretic query expansion is mainly used in probabilistic models that have been extended in different models; Okapi [8] and Divergence From Randomness (DFR) [6][7][9].

The term weighting in the DFR, D_w, is derived by measuring the divergence of the actual term distribution from that obtained under a random process. In this paper we will use Divergence approximation of the binomial (D) normalized with Ratio B of two binomial distributions (B) and the documents length are normalized using normalization 2 (2) [9].

3. N-Order Weight Propagation Technique

In Weight Propagation (WP) [10][11], relevant document propagate positive weights to documents close by in vector similarity space, while documents judged not relevant propagate negative weights to such neighbouring documents and the documents with the highest weights are retrieved and are called the top_set. The N-Order Weight Propagation (NWP) technique [2] makes the documents that received the highest weights propagate positive and negative weights themselves to documents close by in vector similarity space. The documents chosen in the further propagation are selected based on threshold Epsilon ε from the top-set.

NWP showed improvement over WP by propagating one-step further as a 3-Order weight propagation technique or by propagating more steps further as 4-Order, 5-Order and so on. In most cases, improvement becomes stable when NWP reaches 4-Order or 5-Order.

NWP and WP were tested on Web collections that are considered heterogeneous in nature given that these collections contain documents that cover various subjects. However, other type of collections exist which are non-Web and are small to medium in size and enclose documents related to the same subject. These collections are considered homogeneous in nature. Many of the relevance feedback techniques perform well in Web collections but fail in homogeneous collections.

4. Experimental Design

Many standard collections are available for testing an Information Retrieval (IR) system [1]. Commonly used are the Text REtrieval Conference (TREC)[2] small to medium test collections Cranfield, Medline, CISI, LISA and NPL. (Table 1)

Table 1. Test Collections

Test Collection	Size in Megabytes	No of Documents	No of queries	No of Terms	Topics
Medline	1.05	1,033	30	8,915	Medicine
Cranfield	1.40	1,400	225	4,217	Aeronautical engineering
CISI	1.98	1,460	76	5,591	Information Science
NPL	3.02	11,429	93	7,934	Electronic Engineering
LISA	3.04	6,003	35	11,291	Library & Information Science

These collections have very different characteristics in subject matter, record structure, size, mean lengths of documents and queries, mean number of relevant documents per query [12] so that the

[2] Information can be found at http://trec.nist.gov/

experiments will be diverse and fairly complete in assessing relevance feedback techniques. Various models are used in the experiments such as the vector using the tf.idf weighting scheme, and the probabilistic model using the DB2 model in the context of the Divergence from Randomness (DFR) framework[6][7][9]. For our experiments, we have employed the residual collection technique [13] and Simulated Positive, Negative and Pseudo RF [13][14][15]. The effectiveness of the technique is measured using precision at recall level 0.1, precision at recall level 0.3, and the Interpolated Mean Average Precision (MAP) which is the average precision at recall levels=0.0, 0.1… 1.0 [16][17].

5. Experimental Results

The assessments were done as follows: For each query, an initial document ranking was obtained. Ten retrieved documents were taken from the top of the ranked list and used for one iteration of query reformulation. The results of the relevance feedback techniques Rocchio and KL were added to the N-Order Weight Propagation (NWP) and Weight Propagation (WP) results. In addition, comparisons were made between the baseline and both WP and NWP for precision at recall 0.1 using the Student t test over the five test collections in the vector and the probabilistic models.

5.1. Simulated and Pseudo-Relevance Feedback in Vector Model

When tested on small test collections, NWP gave better results than WP starting from low recall levels and beyond. Epsilon ε was set to 1.2 for WP as described by the author to be the best value to optimize the performance. However, it is shown in this paper that decreasing the value of Epsilon ε to 0.8 when using NWP improves the precision to an extent that NWP outperforms WP on all recall levels in simulated and Pseudo relevance feedback.

Medline collection performed better than other collections in Positive and negative feedback whereas NPL gave the best performance in pseudo relevance. The Cranfield collection performed the worst in all relevance feedback an didn't outperformed the baseline at recall 0.1.

Table 2. Experimental Results

Vector Model with tf.idf Weighting Scheme

		Pseudo Relevance Feedback					Positive Relevance Feedback					Negative Relevance Feedback				
	Recall Level	Rocchio	WP	3-Order	4-Order	5-Order	Rocchio	WP	3-Order	4-Order	5-Order	Rocchio	WP	3-Order	4-Order	5-Order
Medline	Prec @1	0.753	0.793	0.783	0.777	0.779	0.804	0.863	0.993	0.979	0.979	0.675	0.702	0.987	0.983	0.588
	Prec @3	0.649	0.637	0.640	0.649	0.649	0.657	0.710	0.915	0.901	0.900	0.491	0.524	0.778	0.767	0.435
	MAP	0.399	0.467	0.486	0.488	0.488	0.399	0.478	0.653	0.646	0.645	0.317	0.425	0.530	0.531	0.369
NPL	Prec @1	0.218	0.195	0.382	0.381	0.381	0.264	0.280	0.280	0.281	0.296	0.241	0.268	0.432	0.485	0.533
	Prec @3	0.133	0.122	0.210	0.210	0.209	0.147	0.106	0.116	0.116	0.119	0.116	0.144	0.175	0.214	0.239
	MAP	0.111	0.105	0.168	0.168	0.168	0.125	0.103	0.111	0.111	0.114	0.102	0.120	0.165	0.189	0.202
Cranfield	Prec @1	0.306	0.320	0.291	0.291	0.291	0.298	0.257	0.277	0.298	0.298	0.298	0.208	0.234	0.225	0.222
	Prec @3	0.119	0.139	0.139	0.139	0.139	0.156	0.123	0.145	0.153	0.153	0.156	0.116	0.147	0.149	0.146
	MAP	0.112	0.112	0.114	0.114	0.114	0.125	0.121	0.129	0.134	0.134	0.125	0.107	0.137	0.136	0.135
LISA	Prec @1	0.151	0.169	0.257	0.279	0.278	0.185	0.276	0.302	0.317	0.317	0.175	0.175	0.235	0.248	0.246
	Prec @3	0.074	0.127	0.134	0.148	0.151	0.101	0.128	0.144	0.144	0.153	0.107	0.118	0.140	0.146	0.149
	MAP	0.067	0.082	0.097	0.104	0.105	0.079	0.082	0.097	0.100	0.100	0.071	0.074	0.093	0.094	0.092
CISI	Prec @1	0.277	0.254	0.352	0.351	0.351	0.334	0.365	0.360	0.355	0.346	0.279	0.317	0.498	0.496	0.499
	Prec @3	0.185	0.179	0.209	0.209	0.208	0.208	0.165	0.169	0.171	0.167	0.169	0.180	0.214	0.214	0.217
	MAP	0.148	0.142	0.178	0.178	0.178	0.178	0.135	0.150	0.152	0.150	0.142	0.148	0.209	0.212	0.214
	Significance P@10		-	∗∗	∗	∗		-	∗∗∗	∗∗∗	∗∗∗		-	∗∗	∗∗	∗∗∗

–	N-order not significantly better than baseline
∗∗∗	N-order significantly better than baseline at 5.0% level
∗∗	N-order significantly better than baseline at 1.0% level
∗	N-order significantly better than baseline at 0.1% level

5.2. Simulated and Pseudo-Relevance Feedback in Probabilistic Model

These experiments are done by comparing the NWP technique against the WP technique in the probabilistic model. The term weighting technique in this section is founded on the DFR model based DB2. Results from the Information-theoretic query expansion Information-theoretic query expansion [11][13] is used as a relevance feedback technique where the topmost documents to be assessed are called “Elite_set T of documents” and the most informative terms are selected by using the information-theoretic Kullback-Leibler divergence function (KL).

A comparison showed that the precision with probabilistic models was better than the precision with vector model at most recall levels and.

WP outperformed the baseline at low recall levels particularly when tested on the Medline test collection, however NWP outperformed WP. NPL performed the best followed by CISI. Cranfield collection performed badly in pseudo relevance feedback but outperformed the baseline in positive and negative feedback

Table 3. Experimental Results

Probabilistic Model using DB2 in the context of DFR

	Recall Level	Pseudo Relevance Feedback					Positive Relevance Feedback					Negative Relevance Feedback				
		Rocchio	WP	3-Order	4-Order	5-Order	Rocchio	WP	3-Order	4-Order	5-Order	Rocchio	WP	3-Order	4-Order	5-Order
Medline	Prec @1	0.673	0.769	0.768	0.770	0.770	0.648	0.770	0.951	0.962	0.962	0.379	0.674	0.614	0.606	0.603
	Prec @3	0.425	0.641	0.678	0.689	0.691	0.442	0.689	0.79	0.804	0.809	0.237	0.573	0.518	0.507	0.506
	MAP	0.290	0.532	0.555	0.554	0.555	0.294	0.554	0.624	0.627	0.628	0.173	0.436	0.399	0.391	0.389
NPL	Prec @1	0.325	0.315	0.495	0.499	0.499	0.340	0.429	0.446	0.432	0.428	0.069	0.335	0.491	0.524	0.564
	Prec @3	0.183	0.193	0.326	0.325	0.325	0.183	0.187	0.196	0.190	0.189	0.043	0.214	0.281	0.279	0.302
	MAP	0.144	0.160	0.246	0.246	0.246	0.147	0.157	0.172	0.169	0.168	0.031	0.162	0.209	0.221	0.232
Cranfield	Prec @1	0.206	0.207	0.131	0.130	0.136	0.130	0.232	0.23	0.230	0.230	0.032	0.169	0.284	0.283	0.283
	Prec @3	0.077	0.083	0.08	0.079	0.080	0.062	0.126	0.124	0.124	0.124	0.026	0.109	0.101	0.099	0.099
	MAP	0.071	0.084	0.075	0.075	0.076	0.053	0.095	0.094	0.093	0.093	0.019	0.092	0.126	0.124	0.123
LISA	Prec @1	0.113	0.191	0.195	0.203	0.203	0.126	0.296	0.296	0.296	0.296	0.009	0.167	0.135	0.124	0.125
	Prec @3	0.049	0.068	0.068	0.068	0.068	0.052	0.074	0.087	0.087	0.087	0.009	0.089	0.069	0.068	0.065
	MAP	0.046	0.062	0.062	0.063	0.063	0.050	0.070	0.071	0.071	0.071	0.006	0.070	0.06	0.058	0.057
CISI	Prec @1	0.246	0.290	0.352	0.352	0.287	0.255	0.302	0.497	0.494	0.343	0.121	0.294	0.285	0.285	0.283
	Prec @3	0.170	0.165	0.216	0.216	0.169	0.171	0.194	0.234	0.229	0.167	0.088	0.152	0.166	0.170	0.171
	MAP	0.134	0.122	0.182	0.182	0.129	0.139	0.155	0.226	0.224	0.149	0.075	0.129	0.133	0.134	0.133
Significance P@10		***	**	**	**			*	*	*	**		*	*	*	*

–	N-order not significantly better than baseline
***	N-order significantly better than baseline at 5.0% level
**	N-order significantly better than baseline at 1.0% level
*	N-order significantly better than baseline at 0.1% level

6. Conclusion

In this paper, we have tested a previously developed relevance technique called N-Order Weight Propagation (NWP) and Weight Propagation (WP) on non-web small to medium test collections that contain mainly documents related to the same subject. The experiments showed that WP and NWP outperformed the baseline by using various models such as the vector using the tf.idf weighting scheme, and the Probabilistic model using the DB2 model in the context of the Divergence from Randomness framework. Many relevance techniques were applied such as the Pseudo, Positive and Negative relevance feedback.

7. References

[1] Baeza-Yates R and Ribeiro-Neto B. Modern information retrieval. New York: Addison-Wesley, page 10, 1999

[2] Yamout F. N-Order Weight Propagation – Information Retrieval 2011 (FG-IR), Germany, 2011

[3] Rocchio G Jr. Relevance Feedback Information Retrieval. Chapter 14. In G. Salton (Ed.), The SMART Retrieval System: Experiments in Automatic Document Processing. Prentice Hall. 1971

[4] Salton G. and McGill M. J. (1983) “Introduction to Modern Information Retrieval”. McGraw Hill, New York, 1983.

[5] Carpineto C., De Mori R., Romano G., and Bigi B., (2001) “An Information Theoretic Approach to Automatic Query Expansion”, ACM Transactions on Information Systems, 19(1):1-27, 2001.

[6] Amati G, Carpineto C, and Romano G. Italian Monolingual Information Retrieval with Prosit. In Proceedings of CLEF (Cross Language Evaluation Forum 2002, Rome, Italy, pages 182-191, 1992

[7] Amati G, Carpineto C, and Romano G. Fub at Trec-10 web track: A Probabilistic Framework for Topic Relevance Term Weighting. In Proceedings of the 10th Text REtrieval Conference (TREC-10), NIST Special Publication 500-250, Gaithersburg, MD, USA pages 182-191, 2001

[8] Robertson S E, Walker S, and Beaulieu M M. Okapi at TREC-7: Automatic ad-hoc, Filtering, VLC and interactive track. In Proceedings of the 7th Text Retrieval Conference (TREC7), Nist Special Publication 500-242, Gaithersburg, MD, USA, pages 253-264, 1998

[9] Amati G and van Rijsbergen C J. Probabilistic Models of Information Retrieval Based on Measuring Divergence from Randomness. ACM Transactions on Information Systems, 20(4), pages 357-389

[10] Yamout F., Oakes M. and Tait J. "Relevance Feedback using Weight Propagation" – 28th European Conference on Information Retrieval ECIR 2006 LNCS 3936– London, UK, pages 575-578

[11] Yamout F., Oakes M. and Tait J. "Relevance Feedback using Weight Propagation Compared with Information-theoretic Query Expansion" – 29th European Conference on Information Retrieval ECIR 2007 LNCS 4425 – Rome, Italy, pages 258-270

[12] Lopez C, Guerrero V. and Moya F. Genetic Algorithms in Releance Feedback: A Second Test and new Contrubutions. Information Processing and Management 39 (2003) 669-687

[13] Ruthven I and Lalmas M. A Survey on the Use of Relevance Feedback for Information Access Systems. Knowledge Engineering Review. Vol. 18. Issue 2, pages 95-145

[14] Buckley, C., Salton, G., and Allan, J. Automatic Retrieval with Locality Information Using Smart. In the First Text REtrieval Conference (TREC-1), National Institute of Standards and Technology, Gaithersburg, MD, 1992, pages 59-72

[15] Efthimiadis, N. E. Query Expansion. In Annual Review of Information Systems and Technology, Vol. 31, 1996, pages 121-187

[16] Kent A., Berry M. Leuhrs F.U. and Perry J.W. Machine Literature Searching VIII: Operational Criteria for Designing Information Retrieval Systems. American documentation, 6(2), 1955, pages 93-101

[17] Saracevic T. Relevance: A Review of and a Framework for the Thinking on the Notion in Information Science. Journal of the American Society for Information Science, November-December, 1975, pages 321-343