A New Model for Probabilistic Information Retrieval on the Web

Gabriel Mateescu
National Research Council
Ottawa Canada
gabriel.mateescu@nrc.ca

Masha Sosonkina
Department of Computer Science
University of Minnesota, Duluth MN
masha@d.umn.edu

Paul Thompson
Thayer School of Engineering
Dartmouth College, Hanover NH
Paul.Thompson@dartmouth.edu

Abstract

Academic research in information retrieval did not make its way into commercial retrieval products until the last 15 years. Early web search engines also made little use of information retrieval research, in part because of significant differences in the retrieval environment on the Web, such as higher transaction volume and much shorter queries. Recently, however, academic research has taken root in search engines. This paper describes recent developments with a probabilistic retrieval model originating prior to the Web, but with features which could lead to effective retrieval on the Web. Just as graph structure algorithms make use of the graph structure of hyper-linking on the Web, which can be considered a form of relevance judgments, the model of this paper suggests how relevance judgments of web searchers, not just web authors, can be taken into account in ranking. This paper also shows how the combination of expert opinion probabilistic information retrieval model can be made computationally efficient through a new derivation of the mean and standard deviation of the model's main probability distribution.

Introduction

Academic research in information retrieval did not make its way into commercial retrieval until the last 15 years when products such as Personal Librarian [Kol81] or, later the ranked retrieval mode of Westlaw, WIN [Wes93] became available. Early web search engines also made little use of information retrieval research, in part because of significant differences in the retrieval environment on the Web. Two main differences from earlier retrieval include higher transaction volume and much shorter queries. More recently, however, academic research has taken root in search engines such as Google [BMPW98].

This paper describes recent developments with a probabilistic retrieval model that originated prior to the Web, but which has features which may lead to effective retrieval on the Web. Just as graph structure algorithms such as [BMPW98, Kle99] make use of the graph structure of hyper-linking on the Web, which can be considered a form of relevance judgments, the model of this paper shows how the relevance judgments of web searchers, not just web authors, can be taken into account in ranking. This paper also shows how the combination of expert opinion probabilistic information retrieval model can be made computationally efficient through a new derivation of the mean and standard deviation of the model's main probability distribution.

Background

The Bayesian combination of expert opinion (CEO) approach to probabilistic information retrieval was first described by Thompson [Tho86, Tho90a, Tho90b]. The CEO model is a generalization of the unified probabilistic retrieval model developed by Robertson, Maron, and Cooper (RMC) [RMC82]. The unified model, called Model 3, was an attempt to combine the models of probabilistic information retrieval developed by Maron and Kuhns [MK60], referred to as Model 1, with the probabilistic retrieval model developed by Robertson and Sparck Jones [RSJ76], van Rijsbergen [vR79], Croft and Harper [CH79], and others, referred to as Model 2. As has been the case with most probabilistic retrieval models, those models were based on the use of point probabilities, rather than on probability distributions. The CEO model, by contrast, provides a probability distribution for a document's being judged relevant by a particular user. In early accounts of the model, it was not shown how the mean and standard deviation, or variance, of these distributions could be computationally implemented. Both the mean and standard deviation of the distributions are needed for the combination process, as well as the ranking of retrieved documents by probability of relevance. This paper shows how the mean and standard deviation of the CEO model's distribution can be computed and how the CEO model can be applied to Web document retrieval.

The unified probabilistic retrieval model, Model 3, was developed so that probabilistic evidence of relevance from the two earlier probabilistic models, Models 1 and 2, could be combined in order to produce a more accurate ranking of documents. As stated in the probability ranking principle [vR79]:

> If a reference retrieval system's response to each request is a ranking of the documents in order of decreasing probability of relevance to the user who submitted the request where the probabilities are estimated as accurately as possible on the basis of whatever data have been made available to the system for this purpose, the overall effectiveness of the system to its user will be the best that is obtainable on the basis of those data.

There were several unresolved issues with the unified model as developed by RMC. Robertson [Rob84] has shown that the term independence assumptions on which the model is based lead to inconsistencies. Also, the unified model did not support relevance feedback. The CEO model was developed to overcome these difficulties, as well as to provide a general probabilistic retrieval model that could combine probabilities from multiple probabilistic retrieval models; not only the two models unified by RMC. In particular, it explored the use of subjective probabilities provided by indexers or searchers [Tho88].

The past decade has also seen much research on the combination of results from multiple retrieval algorithms, representations of text and query, and retrieval systems [Cro00]. The motivation for this research has been provided by several empirical studies showing that different algorithms, representations, and systems provide substantially different, though overlapping, sets of relevant documents [CH79, MKN79, KMT+82, SK88]. This activity has manifested itself both in academic research and in the commercial development of various Web meta-search engines [BKFS95, SE97, Sel99, AM01, MRF01],

Related Research

Combination of models has also been an active area of research in other fields, including statistics [HMRV99, Moe99, Moe00], statistical decision theory [Lin83, CW99, RG01], statistical pattern recognition [XKS92, JDM00], machine learning [LW92, LSCP96, FS97, SS98], and neural networks [JJNH91, JJ94, HSY94, TG99, TG96, Has97, HP98, HPJ99].

Many researchers have applied machine learning techniques to automatic text categorization or clustering, e.g. [LSCP96]. Mathematical techniques new to document retrieval, such as singular value decomposition, or latent semantic indexing, have also been applied [DDF+90]. More recently probabilistic variants of latent semantic indexing have been implemented, as well [PRTV98, Hof99, Hof01].

The Combination of Expert Opinion Model

The CEO model applies Lindley's approach to reconciliation of probability distributions [Lin83] to probabilistic information retrieval (PIR). In this Bayesian model a decision maker with an initial, or prior, probability, or distribution, for some event or parameter θ consults n experts who provide their probabilities are distributions as evidence with which to update his, or her, prior probability distribution via Bayes' theorem to obtain a revised, or posterior, probability, or distribution. In the CEO approach there are two levels of combination. At the upper level the PIR system is considered the decision maker and Models 1 and 2 the experts. At the lower level Models 1 and 2 themselves are derived from CEO. The indexer, or user, in Model 1, or 2, respectively, is seen as a multiple expert - an expert with respect to each use or document property. Each expert, or decision maker, is estimating θ^*, the chance, or long run relative frequency, of success in a Bernoulli sequence of relevance judgments of users for documents. Each Bernoulli sequence is different, but there is a common subsequence which underlies each, so that each expert, or decision maker, can be seen as estimating θ^* for the underlying subsequence. The parameter actually used in the model is θ, where $\theta = \log[\theta^*/(1-\theta^*)]$.

In the CEO model the evidence provided to the decision maker by models being combined are the mean and standard deviations of the experts' distributions. The decision maker's opinion of the experts' expertise, i.e., the weighting of the experts' evidence, is expressed by assuming that $p(m|s,\theta)$ is normal with mean $\alpha + \beta\,\theta$ and standard deviation γs. α, β, and γ are parameters that can be determined through data, i.e., relevance judgments, or a decision maker's subjective belief.

A simplified version of the CEO model was used in the first and second Text REtrieval Conferences (TREC) [Tho93, Tho94]. In this version the mean and standard deviation of the model's main distribution was calculated using approximate techniques. More importantly, relevance feedback was not incorporated in TREC 1. In TREC 2 a form of relevance feedback was used. The ranked retrieval models combined in the TREC 1 system, were weighted by their performance in TREC 1. Unfortunately, due to the many changes made to the models between TREC 1 and TREC 2, the models' performance on TREC 2 was not well predicted by their model 1 performance.

Relevance Feedback

Relevance feedback, i.e., the incorporation of users' judgments as to the relevance of retrieved documents to their information needs, presented a problem with pre-Web retrieval. Laboratory experiments showed that large gains in performance, in terms of precision and recall, could be gained through use of relevance feedback [IS71]. On the other hand, it was assumed that it would not be possible to induce users to provide relevance judgments. Westlaw's WIN was introduced without a relevance feedback capability [Wes93]. By contrast Lexis-Nexis' Freestyle and some web search engines introduced commands that provided relevance feedback based on a single document, rather than a set of relevant documents. These are often called "more like this" commands, where a user selects a single highly relevant returned document and the system returns similar documents. In the TREC conferences and other experimental settings use has been made of pseudo-relevance feedback, where the top n documents are assumed to be relevant and relevance feedback is calculated as though these n documents had actually been judged relevant [SRW01]. As pointed out by Croft et al. [CCTL01] early work on relevance feedback was done with collections of abstracts and results with full text documents have not as good as was anticipated.

In addition to this type of more or less traditional relevance feedback, new forms of relevance feedback have emerged, including implicit relevance feedback, e.g., systems such as Direct Hit [DH] which provide relevance feedback based on mining a user's clickstream, recommender systems [Her01], and rating systems [Del01]. Relevance feedback is usually seen as taking place during a single user's search, but relevance feedback has also been considered in more persistent ways, e.g., in dynamic document spaces [Bra71]. In dynamic document spaces a user's relevance judgments permanently modify the weights of index terms associated with documents.

Probability Distributions used in the Combination of Expert Opinion

As mentioned above, each probabilistic model, e.g., the indexer, or the user, is making an estimate of θ^* for a common underlying Bernoulli subsequence of the overall Bernoulli sequence of viewings of documents by users. Because each model is making these judgments based on the conditioning information available to it, that model's for the sequence's distribution is exchangeable, i.e., the distribution is invariant under finite permutations of its indices [dF74]. A natural distribution to use for a parameter that ranges from 0 to 1, e.g., the proportion of successes in a sequence of relevance judgments, is the beta distribution [Bun84]. It can be very simply updated with each relevance judgment. Graphically the beta distribution can take many shapes, and is thus capable of expressing a wide range of opinion. The CEO algorithm uses a transformation of the beta distribution, the distribution of taken by y, where $y = \log(x/(1-x))$, when x has a beta distribution. It is this distribution, referred to as the *transformed beta distribution*, from which the mean and standard deviation need to be extracted in order to perform the combination of expert opinion and to probabilistically rank retrieved documents.

Computing the Mean and Standard Deviation of the Transformed Beta Distribution

Let y be a continuous random variable whose distribution function is the transformed beta distribution. Then the moment generating function of y is given by ([Tho90b]):

$$\psi(t) = \frac{\Gamma(p+t)\,\Gamma(q-t)}{\Gamma(p)\,\Gamma(q)}$$

where $p, q > 0, -\infty < x < \infty$, and $\Gamma(x)$ is the Gamma function defined by

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt.$$

In this section we derive expressions for the mean and standard deviation of y.

Computing the mean

The formula for the mean μ of y is:

$$\begin{aligned}
\mu &= \frac{d\psi(t)}{dt}\Big|_{t=0} = \frac{d}{dt}\frac{\Gamma(p+t)\Gamma(q-t)}{\Gamma(p)\Gamma(q)}\Big|_{t=0} && (1)\\
&= \frac{d}{dt}\frac{\Gamma(p+t)}{\Gamma(p)}\Big|_{t=0}\ \frac{\Gamma(q-t)}{\Gamma(q)}\Big|_{t=0} + \frac{d}{dt}\frac{\Gamma(q-t)}{\Gamma(q)}\Big|_{t=0}\ \frac{\Gamma(p+t)}{\Gamma(p)}\Big|_{t=0} && (2)\\
&= \frac{\Gamma'(p+t)}{\Gamma(p)}\Big|_{t=0} - \frac{\Gamma'(q-t)}{\Gamma(q)}\Big|_{t=0}, && (3)
\end{aligned}$$

where we have used the facts

$$\frac{dt\ (p+t)}{dt} = 1 \quad \text{and} \quad \frac{dt\ (q-t)}{dt} = -1. \qquad (4)$$

Because $\Gamma'(x \pm t)\Big|_{t=0} = \Gamma'(x)$ we get from (3)

$$\mu = \frac{\Gamma'(p)}{\Gamma(p)} - \frac{\Gamma'(q)}{\Gamma(q)}. \qquad (5)$$

It can be shown (see [WW90]) that

$$\Gamma(x) = \lim_{n\to\infty} \frac{n^x n!}{x(x+1)\cdots(x+n)}. \tag{6}$$

It is easy to show that

$$\frac{n^x n!}{x(x+1)\cdots(x+n)} = e^{x(\ln n - 1 - 1/2 - \cdots - 1/n)} \frac{1}{x} \frac{e^{x/1}}{1+x/1} \frac{e^{x/2}}{1+x/2} \cdots \frac{e^{x/n}}{1+x/n}.$$

Therefore, substituting in (6) we get:

$$\Gamma(x) = e^{-Cx} \frac{1}{x} \prod_{n=1}^{\infty} \frac{e^{x/n}}{1+x/n}, \tag{7}$$

where C is the Euler-Macheroni constant defined by the limit

$$C = \lim_{n\to\infty} \left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} - \ln n\right).$$

With ten decimal places, the value of C is

$$C = 0.5772156649.$$

Taking the logarithm of (7) and differentiating gives:

$$\frac{\Gamma'(x)}{\Gamma(x)} = -C - \frac{1}{x} + \sum_{i=1}^{\infty} \frac{x}{i(x+i)}. \tag{8}$$

We are now ready to compute μ from (5):

$$\begin{aligned} \mu &= -C - \frac{1}{p} + \sum_{i=1}^{\infty} \frac{p}{i(p+i)} - \left(-C - \frac{1}{q} + \sum_{i=1}^{\infty} \frac{q}{i(q+i)}\right) & (9) \\ &= \frac{p-q}{pq} + \sum_{i=1}^{\infty} \frac{p}{i(p+i)} - \sum_{i=1}^{\infty} \frac{q}{i(q+i)}. & (10) \end{aligned}$$

Computing the standard deviation

The standard deviation σ^2 of y is defined as

$$\sigma^2 = E[y^2] - \mu^2$$

where $E[y^2]$ is the second moment of y which is computed using the formula:

$$\begin{aligned} E[y^2] &= \frac{d^2\psi(t)}{dt^2}\Big|_{t=0} = \frac{d^2}{dt^2} \frac{\Gamma(p+t)\Gamma(q-t)}{\Gamma(p)\Gamma(q)}\Big|_{t=0} \\ &= \frac{\Gamma''(p+t)\Gamma(q-t) - 2\Gamma'(p+t)\Gamma'(q-t) + \Gamma(p+t)\Gamma''(q-t)}{\Gamma(p)\Gamma(q)}\Big|_{t=0} \end{aligned}$$

where we have used the facts (4). Thus,

$$E[y^2] = \frac{\Gamma''(p)}{\Gamma(p)} + \frac{\Gamma''(q)}{\Gamma(q)} - 2\,\frac{\Gamma'(p)}{\Gamma(p)}\,\frac{\Gamma'(q)}{\Gamma(q)}.$$

Formally differentiating (8) we get:

$$\frac{\Gamma''(x)}{\Gamma(x)} = \left(\frac{\Gamma'(x)}{\Gamma(x)}\right)^2 + \frac{1}{x^2} + \sum_{i=1}^{\infty} \frac{1}{(x+i)^2}.$$

Then, substituting (8) in the above relation we obtain:

$$\frac{\Gamma''(x)}{\Gamma(x)} = \left(-C - \frac{1}{x} + \sum_{i=1}^{\infty} \frac{x}{i(x+i)}\right)^2 + \frac{1}{x^2} + \sum_{i=1}^{\infty} \frac{1}{(x+i)^2}. \tag{11}$$

The second moment of y is

$$\begin{aligned} E[y^2] \quad = \quad & \left(-C - \frac{1}{p} + \sum_{i=1}^{\infty} \frac{p}{i(p+i)}\right)^2 + \frac{1}{p^2} + \sum_{i=1}^{\infty} \frac{1}{(p+i)^2} \\ & \left(-C - \frac{1}{q} + \sum_{i=1}^{\infty} \frac{q}{i(q+i)}\right)^2 + \frac{1}{q^2} + \sum_{i=1}^{\infty} \frac{1}{(q+i)^2} \\ & -2\left(-C - \frac{1}{p} + \sum_{i=1}^{\infty} \frac{p}{i(p+i)}\right)\left(-C - \frac{1}{q} + \sum_{i=1}^{\infty} \frac{q}{i(q+i)}\right). \end{aligned}$$

Discussion

The CEO model provides a probabilistic framework for combining probabilistic retrieval models. It can be used with subjective probabilities provided either explicitly, or implicitly by users. It can be used both with in the context of a single search and over time. Search on the Web is different in various ways from traditional online document retrieval. Two of the main differences, higher transaction volume and shorter queries, are differences that can be taken advantage of by the CEO model. First, high transaction volumes mean that there are more documents being seen by users from which relevance judgments can be collected. Second, because queries are so much shorter, on average less than 3 words per query as compared to 7 words more typical of traditional online retrieval, it is important to extend the focus of probabilistic models beyond words in documents and queries. As mentioned above algorithms such as HITS [Kle99] or PageRank [BMPW98] extend the focus to hyper-linking. The CEO model shows how this focus could be further extended to user's relevance judgments, whether explicit, or implicit.

The statistical model of the reconciliation of probability distributions, on which the CEO algorithm is based, has seen significant development in recent years, e.g., [RG01]. Related work has been done in machine learning, e.g., on the weighted majority voting algorithm and on boosting [LW92, LSCP96, FS97, SS98], mixture models [JJNH91, JJ94, HP98, Hof99, Hof01, HPJ99, Moe99, Moe00, CH01, MRF01], and Bayesian model averaging [HMRV99]. Text categorization and clustering have become significant application domains for machine learning research. Algorithms such as boosting [SS98], and support vector machines [Joa01] have achieved good results with text categorization.

The focus of these new machine learning and related techniques has been on the document collection, not on the user and the user's information need. As noted by Papadimtriou et al. [PRTV98], "The approach in this body of work [probabilistic information retrieval] is to formulate information retrieval as a problem of learning the concept of 'relevance' that relates documents and queries. The corpus and its correlations plays no central role. In contrast our focus is on the probabilistic properties of the corpus." This focus on the collection ignores the probabilistic evidence provided by an analysis of the user and the user's information need. Relevance is better understood as a relation between the user's information need, which is represented by the query, and the intellectual content of the document, which is represented by the text of the document [Wil73]. While the text of queries and documents may model this latent, deeper structure, especially in the case of the document, user's relevance judgments [CCTL01] and mixed-initiative interaction [HMK99] provide additional evidence of the user's information need. Much research in probabilistic information retrieval is currently focused on language models [Pon98, PC98, CCL01]. Language models are also mainly applied to collections, rather than users, though Lafferty and Zhai [LZ01] provide two language models, one for the document and one for the query, and perform retrieval by measuring the similarity of the two language models.

The CEO model predated much of the research discussed above in the fields of statistics, neural networks, and machine learning. Lindley's [Lin83] model of reconciliation of distributions, now called *Supra-Bayesian pooling* on which the CEO model is based, is still one of the leading theories in the Bayesian approach to combining expert opinions [RG01]. The basic framework of the CEO model appears to be sound, but the model still needs to be completely implemented and empirically tested. In the process of doing so it is likely that the model can be improved through the incorporation of some aspects of the more recent research discussed above. In particular, although there is long-standing precedence in the decision theory literature [Bun84] for using the beta distribution, as discussed above, to model expert opinion, it may be that techniques from Bayesian model averaging [HMRV99] could lead to more accurate modeling. With respect to representation of experts' opinion, the CEO model only requires a mean and standard deviation – not a specific distributional form. More generally, mixture models now being explored in the context of information retrieval, e.g., [Hof99, Hof01, CH01, MRF01], may inform new developments with the CEO model.

Conclusion

The probability ranking principle calls for taking all available evidence into account when probabilistically ranking documents in response to a user's request.

The CEO algorithm provides a formalism for taking all such evidence into account using Bayesian subjective decision theory. The theoretical strength of the CEO algorithm, its ability to easily incorporate relevance judgments and use the judgments to continuously tune its probability estimates, has also been its practical weakness. The success of recommender and similar systems in some domains, e.g., e-commerce, shows that implicit relevance judgments can be effective and may lead to settings where algorithms such as CEO, which rely heavily on relevance judgments, can be effective. Now that an efficient method of calculating the mean and standard deviation of the transformed beta distribution has been derived, the implementation of the CEO model will be facilitated.

References

[AM01] J.A. Aslam and M. Montague. Models for metasearch. In W.B. Croft, D.J. Harper, D. Kraft, and J. Zobel, editors, *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, New Orleans, September 9-13*, pages 276–284. ACM Press, 2001.

[BKFS95] N. Belkin, P. Kantor, E. Fox, and J. Shaw. Combining the evidence of multiple query representations for information retrieval. *Information Processing & Management*, 31(3):431–448, 1995.

[BMPW98] S. Brin, R. Motwani, L. Page, and T. Winograd. What can you do with a web in your pocket? *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, 1998.

[Bra71] T.L. Brauen. Document vector modification. In G. Salton, editor, *Experiments in Automatic Document Processing*, pages 456–484. Prentice-Hall, Englewood Cliffs, 1971.

[Bun84] D.W. Bunn. *Applied decision analysis.* McGraw-Hill, New York, NY, 1984.

[CCL01] J. Callan, B. Croft, and J. Lafferty. *Proceedings of the Workshop on Language Modeling and Information Retrieval, May 31-June 1.* Carnegie Mellon University, http://la.lti.cs.cmu.edu/callan/Workshops/lmir01/WorkshopProcs/index.html, 2001.

[CCTL01] W.B. Croft, S. Cronen-Townsend, and V. Lavrenko. Relevance feedback and personalization: A language modeling perspective. In *Joint DELOS-NSF Workshop on Personalization and Recommender Systems in Digital Libraries, June 18-20, Dublin, Ireland*, pages 18–20, 2001.

[CH79] W. B. Croft and D.J. Harper. Using probabilistic models of document retrieval without relevance information. *Journal of Documentation*, 45(4):285–295, 1979.

[CH01] D. Cohn and T. Hofmann. The missing link: A probabilistic model of document content and hypertext connectivity. In *Advances in Neural Information Processing Systems (NIPS*13)*. MIT Press, Cambridge, MA, 2001.

[Cro00] W.B. Croft. Combining approaches to information retrieval. In W.B. Croft, editor, *Advances in Information Retrieval: Recent Research from the Center for Intelligent Information Retrieval*, chapter 1, pages 1–36. Kluwer Academic, Boston, MA, 2000.

[CW99] R.T. Clemen and R.L. Winkler. Combining probability distributions from experts in risk analysis. *Risk Analysis*, 19:187–203, 1999.

[DDF$^+$90] S. Deerwester, S.T. Dumais, G.W. Furnas, T.K. Landauer, and R. Harshman. Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6):381–407, 1990.

[Del01] C. Dellarocas. Building trust on-line: The design of reliable reputation reporting mechanisms for online trading communities paper 101. *Center for eBusiness@MIT*, 2001.

[dF74] B. de Finetti. *Theory of Probability: A critical introductory treatment*, volume 1. John Wiley, New York, 1974.

[DH] Direct Hit. *http://www.directhit.com.*

[FS97] Y. Freund and R.E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997.

[Has97] S. Hashem. Algorithms for optimal linear combinations of neural networks. In *International Conference on Neural Networks*, volume 1, pages 242–247, 1997.

[Her01] J. Herlocker. *2001 ACM SIGIR Workshop on Recommender Systems Notes.* ACM Press, http://www.cs.orst.edu/∼herlock/ rsw2001/workshop_notes.html, 2001.

[HMK99] S. Haller, S. McRoy, , and A. Kobsa. *Computational Models of Mixed-Initiative Interaction.* Kluwer, Boston MA, 1999.

[HMRV99] J.A. Hoeting, D. Madigan, A.E. Raftery, and C.T. Volinsky. Bayesian model averaging: A tutorial. *Statistical Science*, 14(4):382–417, 1999.

[Hof99] T. Hofmann. Probabilistic latent semantic indexing. In M. Hearst, F. Harper, Gey, and R. Tong, editors, *Proceedings of SIGIR'99 22nd International Conference on Research and Development in Information Retrieval, New Orleans, September 9-13*, pages 50–57, 1999.

[Hof01] T. Hofmann. What people (don't) want. In *European Conference on Machine Learning (ECML)*, 2001.

[HP98] T. Hofmann and J. Puzhica. Unsupervised learning from dyadic data. Technical Report TR-98-042, University of California, Berkeley, 1998.

[HPJ99] T. Hofmann, J. Puzhica, and M. Jordan. Learning from dyadic data. In M.S. Kearns, S.A. Solla, and D. Cohn, editors, *Advances in Neural Information Processing Systems*, number 11. MIT Press, Cambridge, MA, 1999.

[HSY94] S. Hashem, B. Schmeiser, and Y. Yih. Optimal linear combinations of neural networks: An overview. In *IEEE World Congress on Computational Intelligence, IEEE International Conference on Neural Networks*, volume 3, pages 1507–1512, 1994.

[IS71] E. Ide and G. Salton. Interactive search strategies and dynamic file organization in information retrieval. In G. Salton, editor, *Experiments in Automatic Document Processing*, pages 373–393. Prentice-Hall, Englewood Cliffs, NJ, 1971.

[JDM00] A.K. Jain, R.P.W. Duin, and J. Mao. Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):4–37, 2000.

[JJ94] M.I. Jordan and R.A. Jacobs. Hierarchical mixtures of experts and the em algorithm. *Neural Computation*, 6:181–214, 1994.

[JJNH91] R.A Jacobs, M.I. Jordan, S. Nowlan, and G.E. Hinton. Adaptive mixtures of local experts. *Neural Computation*, 3:1–12, 1991.

[Joa01] T. Joachim. A statistical learning model of text categorization for support vector machines. In W.B. Croft, D.J. Harper, D. Kraft, , and J. Zobel, editors, *SIGIR 2001 Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, New Orleans, September 9-13*, pages 128–136, 2001.

[Kle99] J.M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–622, 1999.

[KMT+82] J. Katzer, M. McGill, J. Tessier, W. Frakes, and P. Das-Gupta. A study of the overlap among document representations information technology. *Research and Development*, 1(4):261–274, 1982.

[Kol81] M.B. Koll. Information retrieval theory and design based on a model of the user's concept relations. In R.N. Oddy, S.E. Robertson, C.J. van Rijsbergen, and P.W. Williams, editors, *Information Retrieval Research*. Butterworths, Boston, MA, 1981.

[Lin83] D.V. Lindley. Reconciliation of probability distributions. *Operations Research*, 31(5):866–880, 1983.

[LSCP96] D.D. Lewis, R.E. Schapire, J.P. Callan, and R. Papka. Training algorithms for linear text classifiers. In H.-P. Frei, D. Harman, P. Schauble, , and R. Wilkinson, editors, *SIGIR '96: Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Konstanz*, pages 298–306. Hartung-Gorre Verlag, 1996.

[LW92] N. Littlestone and M. Warmuth. The weighted majority voting algorithm. Technical Report UCSC-CRL-91-28, University of California, Santa Cruz, 1992.

[LZ01] J. Lafferty and C. Zhai. Document language models, query models, and risk minimization for document retrieval. In W.B. Croft, D.J. Harper, D. Kraft, , and J. Zobel, editors, *SIGIR 2001 Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, New Orleans, September 9-13*, pages 111–119, 2001.

[MK60] M.E. Maron and J.L. Kuhns. On relevance, probabilistic indexing and information retrieval. *Journal of the ACM*, 7(3):216–244, 1960.

[MKN79] M. McGill, M. Koll, and T. Noreault. An evaluation of factors affecting document ranking by information retrieval systems. Final report for grant nsf-ist-78-10454 to the national science foundation, Syracuse University, 1979.

[Moe99] P. Moerland. A comparison of mixture models for density estimation. In *Proceedings of the International Conference on Artificial Neural Networks (ICANN'99)*, 1999.

[Moe00] P. Moerland. Mixtures of latent variable models for density estimation and classification. Research Report IDIAP-RR 00-25, IDIAP, 2000.

[MRF01] R. Manmatha, T. Rath, and F. Feng. Modeling score distributions for combining the outputs of search engines. In W.B. Croft, D.J. Harper, D. Kraft, and J. Zobel, editors, *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, New Orleans, September 9-13*, pages 267–275. ACM Press, 2001.

[PC98] J. Ponte and W.B. Croft. A language modeling approach to information retrieval. In W.B. Croft, A. Moffat, C.J. van Rijsbergen, R. Wilkinson, and J. Zobel, editors, *SIGIR '98 Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Melbourne, Australia August 24-28*, pages 275–281. York Press, 1998.

[Pon98] J. Ponte. *A language modeling approach to information retrieval.* PhD thesis, University of Massachusetts, Amherst, 1998.

[PRTV98] C.H. Papadimitriou, P. Raghavan, H. Tamaki, and S. Vempala. Latent semantic indexing: a probabilistic analysis. In *Proceedings of the seventeenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems*, volume 3, pages 159–168, 1998.

[RG01] P.J. Roback and G.H. Givens. Supra-bayesian pooling of priors linked by a deterministic simulation model. *Communications in Statistics-Simulation and Computation*, 30:447–476, 2001.

[RMC82] S.E. Robertson, M.E. Maron, and W.S. Cooper. Probability of relevance: A unification of two competing models for document retrieval. *Information Technology: Research and Development*, 1(1):1–21, 1982.

[Rob84] S.E. Robertson. Consistencey of the RMC model. Personal communication, 1984.

[RSJ76] S.E. Robertson and K. Sparck Jones. Relevance weighting of search terms. *Journal of the American Society for Information Science*, 27(3):129–146, 1976.

[SE97] E. Selberg and O. Etzioni. The MetaCrawler architecture for resource aggregation on the Web. *IEEE Expert*, 12(1):8–14, 1997.

[Sel99] E.W. Selberg. *Towards Comprehensive Web Search.* PhD thesis, University of Washington, 1999.

[SK88] T. Saracevic and P. Kantor. A study of information seeking and retrieving. part iii. searchers, searches, overlap. *Journal of the American Society for Information Science and Technology*, 39(3):197–216, 1988.

[SRW01] T. Sakai, S.E. Robertson, and S. Walker. Flexible relevance feedback for NTCIR-2. In *Proceedings of the Second NTCIR Workshop Meeting on Evaluation of Chinese & Japanese Text Retrieval and Text Summarization Tokyo, Japan, March 7 - 9, Tokyo.* National Institute of Informatics, 2001.

[SS98] R.E. Schapire and Y. Singer. Improved boosting algorithms using confidence-rated predictions. In *Proceedings of the Eleventh Annual Conference on Computational Learning Theory, Madison, Wisconsin July 24-26*, 1998.

[TG96] K. Tumer and J. Ghosh. Theoretical foundations of linear and order statistics combiners for neural pattern classifiers. Technical Report TR-95-02-98, The University of Texas at Austin, 1996.

[TG99] K. Tumer and J. Ghosh. Linear and order statistics combiners for pattern classification. In A. Sharkey, editor, *Combining Artificial Neural Networks*, pages 127–162. Springer-Verlag, New York, NY, 1999.

[Tho86] P. Thompson. *Subjective probability, combination of expert opinion, and probabilistic approaches to information retrieval*. PhD thesis, University of California, Berkeley, 1986.

[Tho88] P. Thompson. Subjective probability and information retrieval: A review of the psychological literature. *Journal of Documentation*, 44(2):119–143, 1988.

[Tho90a] P. Thompson. A combination of expert opinion approach to probabilistic information retrieval, part 1: The conceptual model. *Information Processing & Management*, 26(3):371–382, 1990.

[Tho90b] P. Thompson. A combination of expert opinion approach to probabilistic information retrieval, part 2: mathematical treatment of ceo model 3. *Information Processing & Management*, 26(3):383–394, 1990.

[Tho93] P. Thompson. Description of the PRC CEO algorithm for TREC. In Harman D.K., editor, *The First Text REtrieval Conference (TREC-1)*, volume National Institute of Standards and Technology Special Publication 500-207, pages 337–342. U.S. Government Printing Office, Washington, D.C., 1993.

[Tho94] P. Thompson. Description of the PRC CEO algorithm for TREC-2. In Harman D.K., editor, *The Second Text REtrieval Conference (TREC-2)*, volume NIST Special Publication 500-215, pages 271–274. U.S. Government Printing Office, Washington, D.C., 1994.

[vR79] C.J. van Rijsbergen. *Information Retrieval*. Buttersworth, London, UK, 2 edition, 1979.

[Wes93] West Publishing Company, St. Paul, MN. *WESTLAW Reference Manual*, 5 edition, 1993.

[Wil73] P. Wilson. Situational relevance. *Information Storage and Retrieval*, 9(7):457–471, 1973.

[WW90] E.T. Whittaker and G.N. Watson. *A Course in Modern Analysis*. Cambridge University Press, Cambridge, UK, 4 edition, 1990.

[XKS92] L. Xu, A. Krzyzak, and C.Y. Suen. Methods of combining classifiers and their applications to handwriting recognition. *IEEE Transactions on Systems, Man, and Cybernetics*, 22(3):418–435, 1992.