8. Training and Testing:

The Predictive Power of the Two Parametric Analysis Methods and One Non-Parametric Analysis Method

8.1 Testing Data

In TREC 5, there were twenty six IR schemes participating in the routing task. Twenty three of them used all the collection documents available, the rest of them only used a subset of the collection (Voorhees & Harman, 1997). We use the output lists of these twenty three schemes as testing data.

There were fifty topics in TREC 5 routing task, therefore we have in all 23×50 = 1,150 output lists. Each output list contains 1,000 documents in ranked order of assigned relevancy scores. From these 1,150 lists, there are ½(23×22×50) = 16,250 pairs for data fusion. As it turns out, for topics 68, 125, 237, 240 and 243, there are no relevant documents in the testing collection. That means, for these five topic no matter how good or how bad a data fusion rule is, there will not be any decrease or increase in precision at the 100th document. We eliminated these five topics from my testing data, so finally we have ½(23×22×45) = 11,385 cases.

For each pair of information retrieval schemes (S_1 , S_2):

1. We divided the precision at the 100th document of the comparatively poorer scheme by the precision at the 100th document of the better scheme to get the ratio of precisions (***r***). If the two schemes have the same precision at the 100th document, the precision ratio is one;
2. We compute the normalized dissimilarity (***z***) between the two IR schemes using the ranked order of all the 1,000 documents of each list);
3. We rank the documents retrieved by the two IR schemes according to the sum of normalized relevancy scores and keep the top one thousand documents,
4. We compare the top 100 documents of the fused list with the list of elevant documents provided by TREC 5 to get the precision at the 100th document, of the fused list, i.e., $P(S_1\ f\ S_2)$, where ***f*** represents the simple symmetric fusion rule.
5. We calculate the relative improvement of the data fusion against an oracle, $E\ (S_1\ f\ S_2)$, by the formula:

$$E\ (S_1 f\ S_2) = \frac{P(S_1 f\ S_2) - \max\{P(S_1), P(S_2)\}}{\max\{P(S_1), P(S_2)\}}$$

8.2 Predictive Power of Two Parametric Methods: Discriminant Function and Logistic Regression Equation

We apply the discriminant function estimated from the training data set to predict the discriminant scores of the cases in the testing data set, and then use the predicted discriminant scores to plot an ROC curve. We also apply the logistic regression equation estimated from the training data set to the testing data set to predict the probability that the effectiveness of data fusion is greater than zero. In this case, we use the predicted probability to plot an ROC curve.

Figure 8.1 shows the ROC curves using the two predicting methods. As in the training data set, the two ROC curves almost completely overlap with each other and are visually not distinguishable. That means that the predictive power of these two methods are practically the same.

From *Figure 8.1*, we can see that when the detection rate is below about 75%, the predictive power of either of these two ROC curves is much better than predicting without any information about normalized dissimilarity and precision ratio.

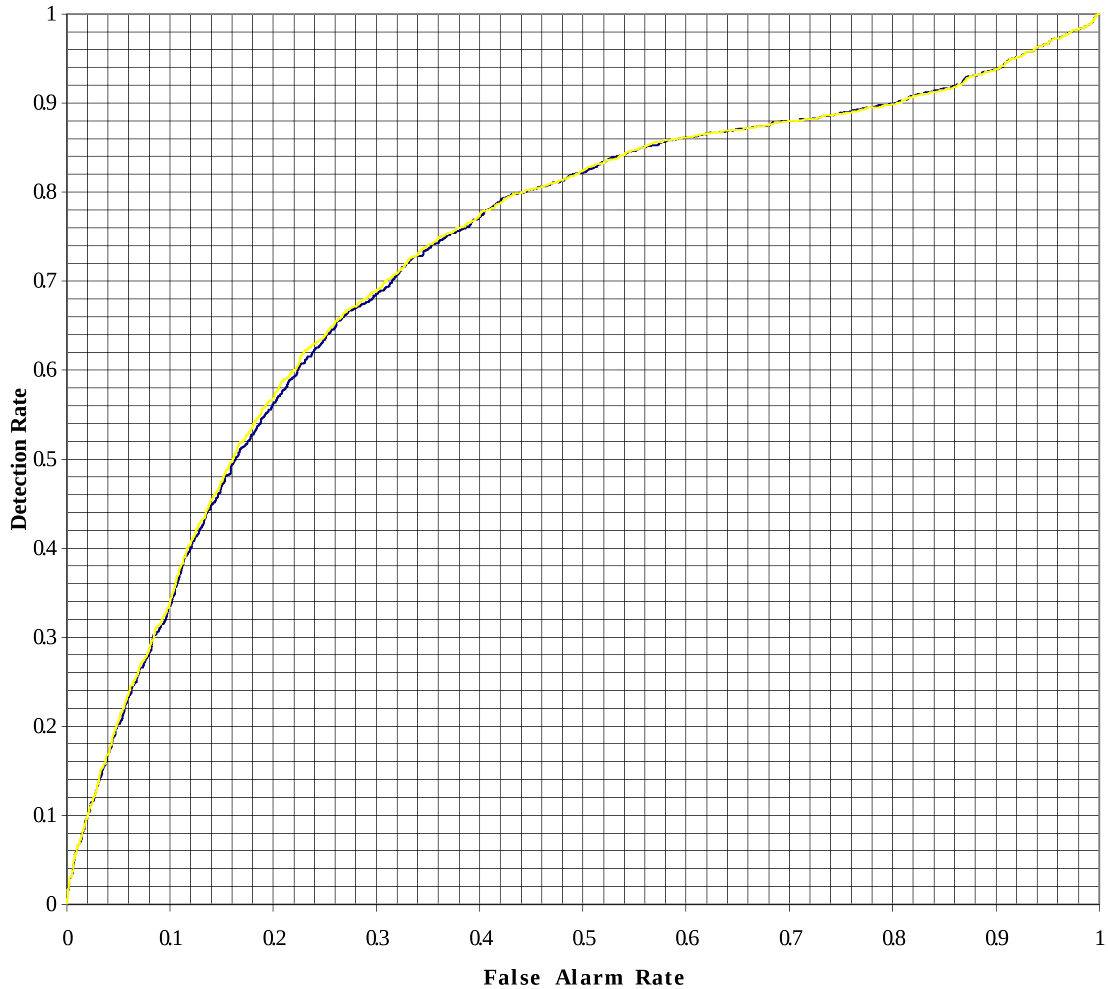

Figure 8.1 *The two ROC curves (testing data set) of predicted discriminant scores by discrminant analysis (dark curve) and predicted probabilities by logistic regression (light curve). They almost completely overlap with each other.*

When the detection rate is below about 75%, it is about one to two times higher than the false alarm rate. For example, when the detection rate is about 60%, the false alarm rate is about 22%, the detection rate is about 2.73 times of the false alarm rate; when the detection rate is about 70%, the false alarm rate is about 31%, the detection rate is about 2.26 times of the false alarm rate.

However, when the detection rate is higher than 75%, although the detection rate is not much higher than the false alarm rate as before, it is still better than chance alone. For example, when the detection rate is about 80%, the false alarm rate is about 47%; when the detection rate is about 90%, the false alarm rate is about 80%, still lower than detection rate.

We see that the two statistical analysis techniques have good predictive power in the low false alarm rate and moderate detection rate region, that is, from about 16% to 31% of false alarm rate and 50% to 70% detection rate.

8.3 Non-Parametric Training and Testing

The above two statistical techniques are parametric techniques. In other words, the model has a few parameters (the Betas). Also, the underlying statistical measures assumes that variables have multivariate normal distribution.

Besides parametric techniques, we can apply a non-parametric analysis which is totally empirical and highly nonlinear.

We group the positive cases and the negative cases of the training data into 100 square bins by defining 10 ranges of normalized dissimilarity and 10 ranges of ratio of precisions. Each bin contains the number of cases in a particular range of normalized dissimilarity and ratio of precisions.

If we divide the number of positive cases in a bin by the number of negative cases in the bin, we build the *Table 8.1* If we arrange the bins in rank order, with the cell have the highest rate of positive cases to negative case ratio ranked as “1”, we get *Table 8.2* (next page).

		Normalized Dissimilarity									
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Ratio of Precisions	0.1	0.000	0.000	0.000	0.504	0.090	0.113	0.280	0.087	0.132	0.221
	0.2	0.000	0.000	0.315	0.000	0.000	0.107	0.031	0.238	0.188	0.793
	0.3	0.000	0.000	0.105	0.360	0.320	0.303	0.182	0.266	0.244	1.222
	0.4	0.000	0.000	0.063	0.291	0.233	0.187	0.343	0.372	0.744	1.453
	0.5	0.000	0.280	0.000	0.170	0.240	0.423	0.460	0.639	1.471	3.393
	0.6	0.000	0.373	0.111	0.123	0.236	0.760	0.667	1.735	0.802	1.512
	0.7	0.000	0.000	0.089	0.161	0.488	0.602	1.220	1.704	1.557	3.664
	0.8	0.000	0.450	0.455	0.686	1.169	1.812	2.262	2.465	2.927	2.240
	0.9	0.000	0.510	0.904	1.492	3.077	3.720	4.217	5.880	6.298	7.132
	1.0	0.734	2.771	4.644	6.184	7.892	8.277	8.860	11.024	14.092	16.348

Table 8.1: Table of comparative performance of data fusion in different ranges of normalized dissimilarity and ratio of precisions. Each cell contains the ratio of positive cases to negative cases in a particular range of normalized dissimilarity and precision ratio.

The highest rank in *Table 8.2* is 1 and the lowest rank is 83. If we use these data to plot an ROC curve, there are 83 points in the curve, point 1 represents the detection rate and false alarm rate we get if we use rank 1 as the cutoff point, point

2 represents the detection rate and false alarm rate we will get if we use rank 2 as cutoff point, etc.

		Normalized Dissimilarity									
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Ratio of Precisions	0.1	83	83	83	44	78	74	58	80	72	66
	0.2	83	83	55	83	83	76	82	63	67	35
	0.3	83	83	77	52	54	56	69	60	61	30
	0.4	83	83	81	57	65	68	53	51	37	29
	0.5	83	59	83	70	62	49	46	41	28	15
	0.6	83	50	75	73	64	36	40	23	34	26
	0.7	83	83	79	71	45	42	31	24	25	14
	0.8	83	48	47	39	32	22	20	19	17	21
	0.9	83	43	33	27	16	13	12	10	8	7
	1.0	38	18	11	9	6	5	4	3	2	1

Table 8.2: Each cell contains the rank order of the comparative performance of a particular range of normalized dissimilarity and precision ratio. The cell with the rank one represents the range of normalized dissimilarity and precision ratio which will give the highest ratio of positive cases to negative cases.

The detection rate at point 1 will be the number of positive cases in the rank 1 bin divided by the total number of positive cases in the data set; the false alarm rate of point 1 will be the total number of negative cases in the rank 1 bin, divided by the total number of negative cases in the data set. The detection rate at point 2 is the number positive cases in the rank 1 bin and rank 2 bin combined, divided by the total number of positive cases in the data set; the false alarm rate of point 2 is the number of negative cases in the rank 1 bin and rank 2 bin combined, divided by the total number of negative cases in the data set. The detection rate of point ***n*** will be the number of positive cases in the rank 1 bin, rank 2 bin, …, rank ***n*** bin combined, divided by the total number of positive cases in the data set. The false alarm rate of point ***n*** is the total number of negative cases in the rank 1 bin, rank 2 bin, …, rank ***n*** bin combined, divided by the total number of cases in the data set.

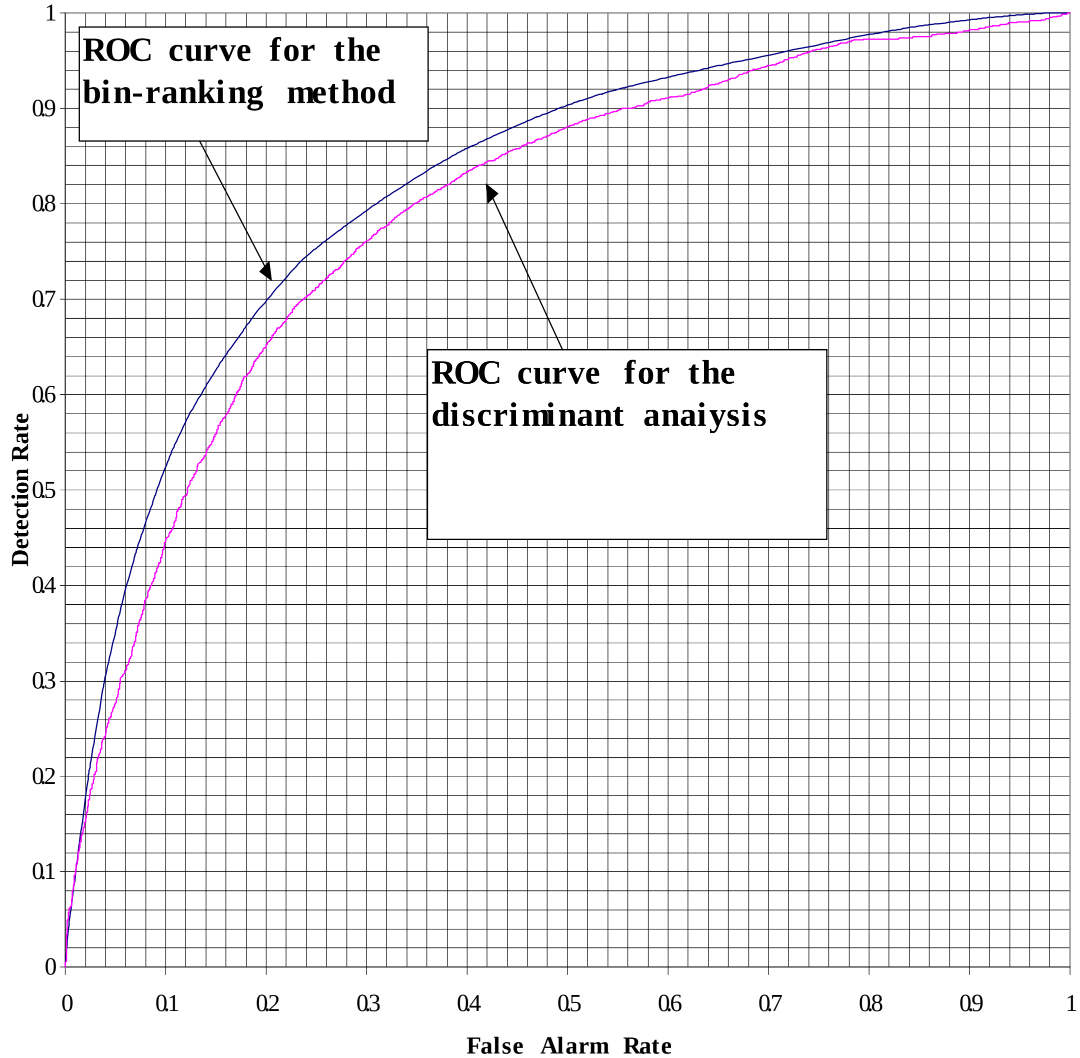

Figure 8.2: Comparison of performance between the non-parametric bin-ranking method and the parametric discriminant analysis for the training data set.

Figure 8.2 (previous page) is the ROC curve of this non-parametric bin-ranking method and the ROC curve of one of the two parametric methods, i.e., linear discrminant analysis.

From *Figure 8.2*, we can see that the ROC curve of this non-parametric method is more powerful than the ROC curves of the parametric method. With the same false alarm rate, the detection rate of the non-parametric method is always higher than that of the discriminant analysis.

For example, when the false alarm rate of the non-parametric method is about 30%, the detection rate is about 79%; while for ROC of the discriminant analysis is about 12%, the detection rate is about 57%, while for the discriminant analysis, when the false alarm rate is about 12%, the detection rate is about 50%.

Since the ROC curve of the bin ranking method is apparently better the parametric method, perhaps rank the bins of the testing data set according the rank order determined by the training set, we will get a better prediction.

We group the cases of the testing data set into the same 100 square bins. Next we divide the number of positive cases in a bin by the number of negative cases in the corresponding bin. We rank the cells according to the rank order of the training data, and then plot an ROC curve accordingly (*Figure 8.3*, next page).

From *Figure 8.3*, we see that the ROC curve of arranging the bins of the testing data set according to the order determined by the training data set is not always concave, as was the ROC curve of the training data set. For example, check the point when the detection rate is about 60% and the false alarm rate is about 21%. These turning points indicate that the bins corresponding at the consecutive points are out of rank order.

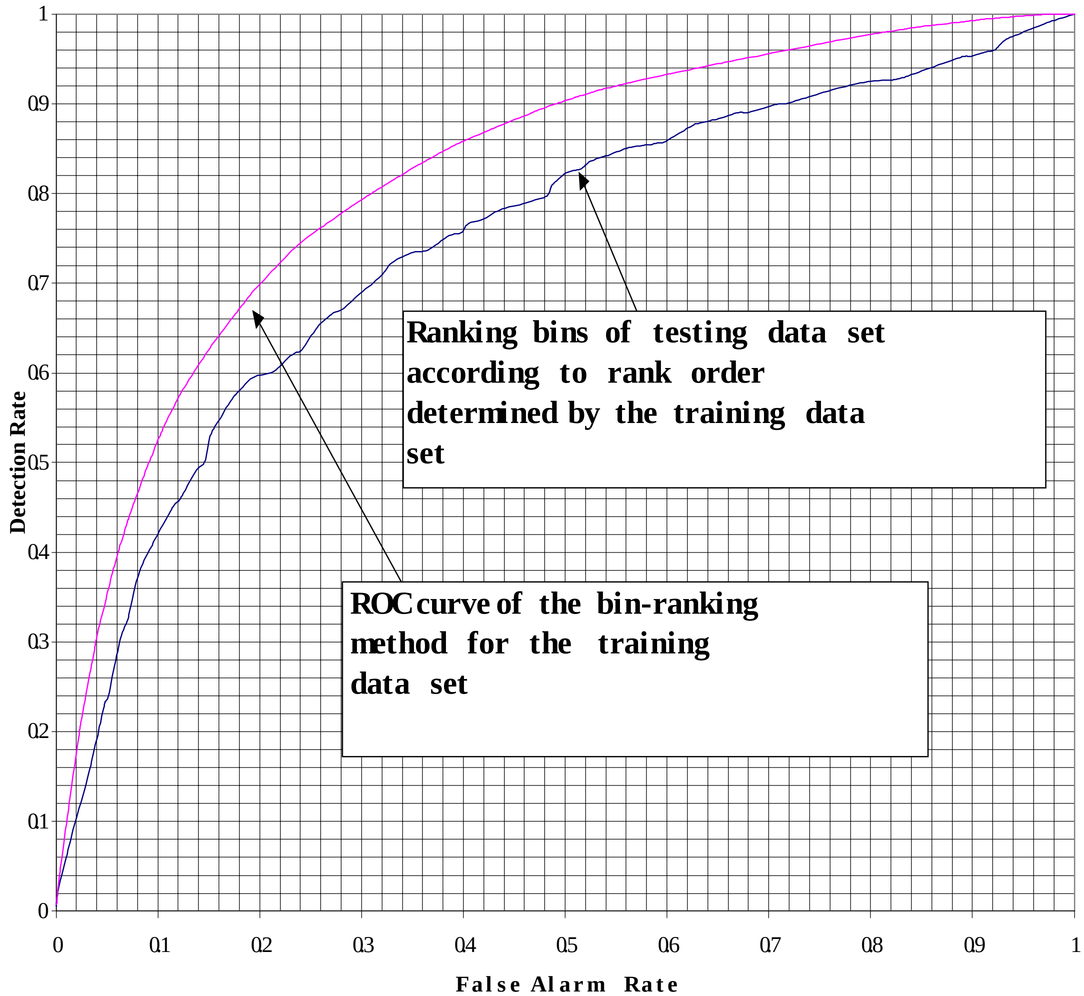

Figure 8.3: The ROC curves of the bin-ranking method: training and testing.

Let's use the point when the detection rate is about 60% and the false alarm rate is about 21% as example. It corresponds to the bin with rank order 23, the exact detection rate is 60.17% and the exact false alarm rate is 21.40%. The ratio of positive cases to negative cases of this rank is 0.1477. The ratio of positive cases to negative cases of the rank 22 bin is 0.2518, while for the rank 24 bin, it is 0.5192 (*Table 8.3.*)

Predicted rank	**Ratio of positive cases to negative cases in the bin**
22 (A)	0.2518
23 (B)	0.1477
24 (C)	0.5192
Predicted rank order: A > B > C	
Ideal rank order: C > A >B	

Table 8.3: predicted rank order vs. ideal rank order for three bins in the testing data set.

From *Figure 8.2* and *Figure 8.3*, it seems that although the bin ranking method is more powerful than the two parametric methods in classifying the training cases into positive and negative cases, its predictive power drops a lot when applied to the testing data set. Of course it is unlikely that applying parameters estimated from training data to testing data will give the same performance. However, the difference between the training stage and testing stage seems to be quite large.

For example, from *Figure 8.3*, at point where the false alarm rate is about 20%, the detection rate of the training data set is about 70% but the detecting rate of the testing data set is about 60% , decreases 10%; when the false alarm rate is about 40%, the detection rate of the training data set is about 86%, the detection rate of the training data set is about 76%, also decrease by 10%.

The ROC curves for discriminant analysis do not change so much from the training data set to the testing data set (*Figure 8.4.)*

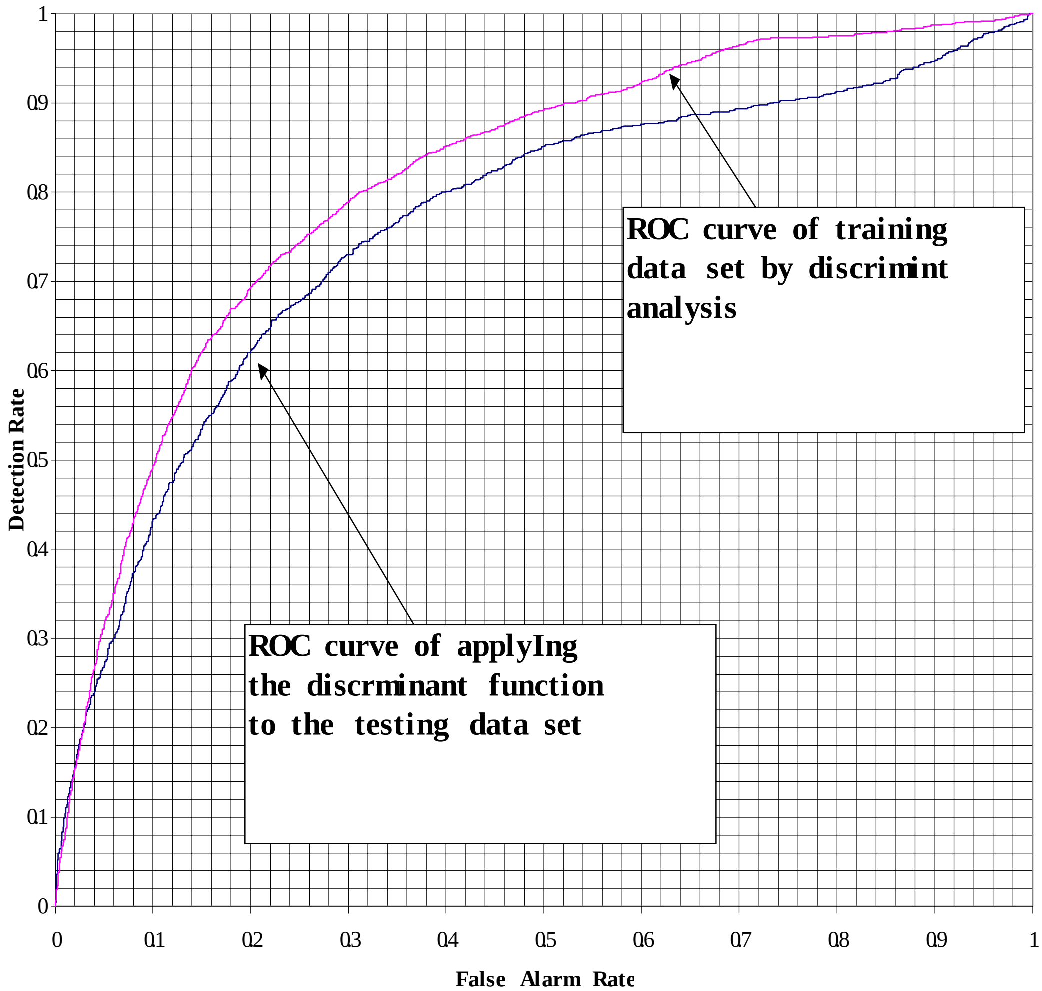

Figure 8.4: The performance of discriminant analysis on the training data set (lighter curve) and the performance of the same discriminant function applied to the testing data set (darker curve) .

From *Figure 8.4*, we see that generally the vertical distance between the two curves is not as great as for the non-parametric method (*Figure 8.3*). The biggest difference is in the range of false alarm rate about 68% to 78%. The ROC curve for the training data suddenly rises at this point while the ROC curve for the testing data suddenly flattens, indicating that it may be a region of over-training. In other words, because of the pattern of the distribution of positive cases and negative cases in the plane of ratio of precisions and normalized dissimilarity, the detection rate suddenly has a more rapid rise than the false alarm rate in this region for this data set, however this is only a specific characteristic of the training data, not a general characteristic for all data set. The rank order of the same region in another data set probably will be different from this training data set. Therefore, applying the trained rank order (which has a sudden improvement) of that region to other data set probably will not get the same rapid rise in detection rate, but a rapid rise in false alarm rate because the actual rank other is different.

8.4 Optimal ROC Curve of the Testing Data Set

We can investigate why there is a comparatively large drop of detection rate when using the non-parametric method, by comparing the ROC curves of the testing data set with the corresponding optimal ROC curve. From *Figure 8.5 (*next page*)*, we can see the predicted ROC curve based on arranging the 100 bins of the testing data set according to the rank order of the bins in the training data set is very close to its corresponding optimal ROC curve which is based on the actual rank order of the bins of the testing data set.

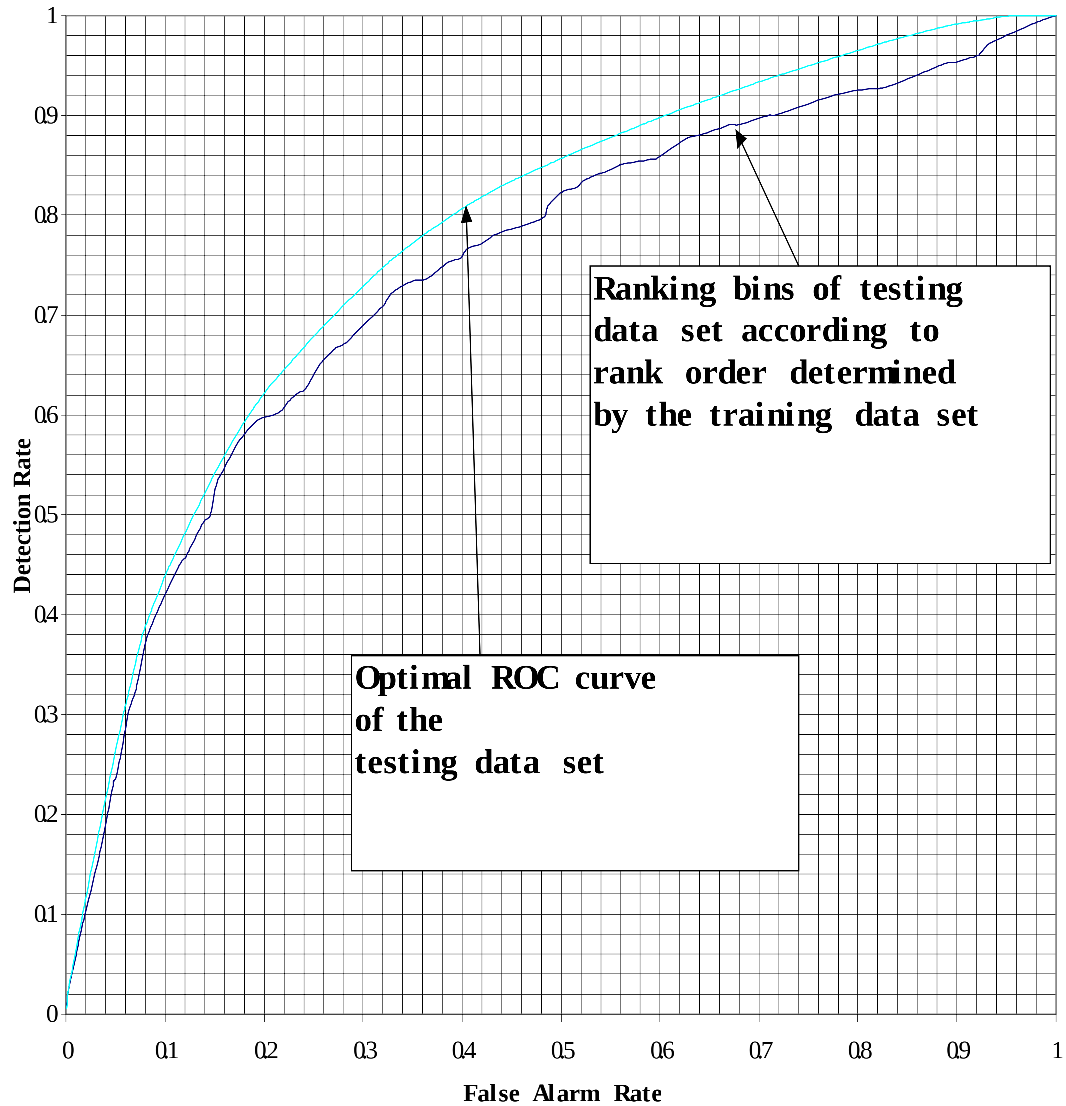

Figure 8.5 Comparing the ROC curves of the training data set to its corresponding optimal ROC curve .

In other words, the predicted ROC curve of the non-parametric method is very close to its corresponding optimal curve. The drop in detection rate from training data set to testing data set is only because the corresponding optimal curve for the testing data set is lower than that of the training data set (*Figure 8.6*).

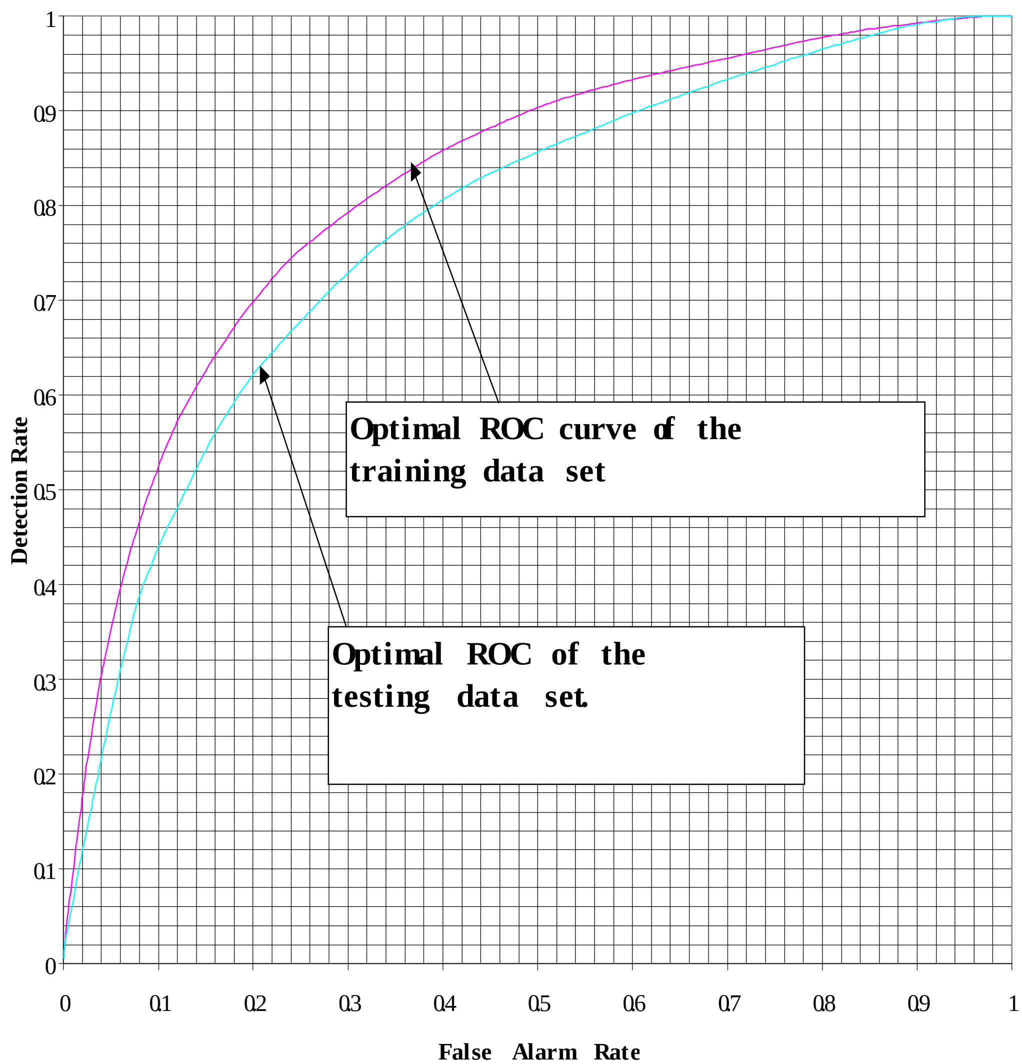

Figure 8.6: *Comparison of the ROC curves for testing and training data set arranging bins according to their actual rank in the data set.*

8.5 The Performance of Non-Parametric Method.

Although compared to the parametric method, e.g., discriminant analysis, the predictive power of the non-parametric bins ranking methods seems to drop more from training data set to testing data set, it may still have a better performance.

Since the ROC curves of the discriminant analysis and logistic regression almost completely overlap each other (see *figure 8.1*), the discussion here also applies to the logistic regression analysis.

Figure 8.7 (next page) is the ROC curves of the two predictive methods: parametric discriminant analysis and the non-parametric bin ranking method. When the false rate is below about 60%, the two ROC curves are very close to each other, with less than 1% difference in detection rate, indicating that the predictive power of the parametric method and the non-parametric method are more or less the same until false alarm rate is about 60%. When the false alarm rate is above 60%, the ROC curve of the non-parametric method is always better than the parametric method by about 2%.

In the above we have compared the performance of different predictive methods using ROC curves. In the following, we are going to discuss how to rationally predict the sign of effectiveness of data fusion rational based on the ROC curve.

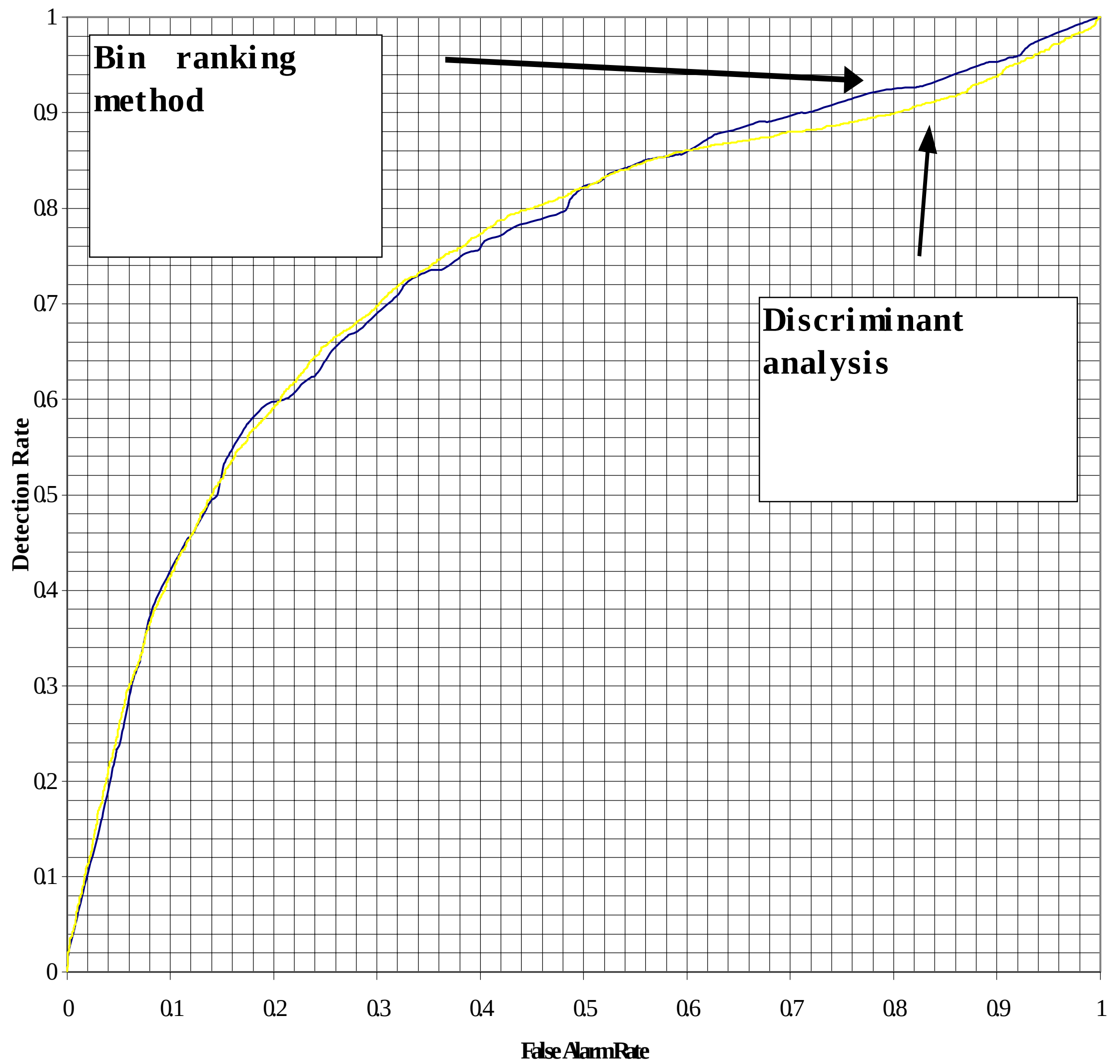

Figure 8.7: *The prediction performance of the two methods in the testing data set: parametric linear discriminant analysis and non-parametric bin ranking method.*

8.6 ROC and Effectiveness Prediction

We represent the problem of detecting positive effectiveness of data fusion in terms of signal detection. In particular, we use the ROC curve of signal detection theory. In the following, we will use the ROC of discriminant analysis as example, to discuss how to use the ROC curves to predict effectiveness of data fusion in IR.

Although we will use only discriminant analysis as example, the discussion also can also be applied to other predictive methods which will assign some kind of descriminant scores to the cases, e.g., probability based on logistic regression analysis, or inverse rank order based on the bin-ranking method.

We have two events, i.e.,, $e_{positive}$ and $e_{negative}$. The former means the effectiveness of data fusion is greater than zero while the latter means the effectiveness of data fusion is less than zero. We also have two responses, i.e.,, $r_{positive}$ and $r_{negative}$: predict as positive and predict as negative respectively. There are four event-response combinations: $(e_{positive}, r_{positive})$, $(e_{positive}, r_{negative})$, $(e_{negative}, r_{positive})$, $(e_{negative}, r_{negative})$, the conditional probabilities of these events satisfy the constrains:

$$Pr(e_{negative}) + Pr(e_{positive}) = 1$$

$$Pr(r_{positive} \mid e_{positive}) + Pr(r_{negative} \mid e_{positive}) = 1$$

$$Pr(r_{positive} \mid e_{negative}) + Pr(r_{negative} \mid e_{negative}) = 1$$

We can represent the above equations in *Table 11.4.*

	$r_{positive}$	$r_{negative}$	**Sum of Pr**
$e_{positive}$	$Pr(r_{positive} \mid e_{positive})$	$Pr(r_{negative} \mid e_{positive})$	*1*
$e_{negative}$	$Pr(r_{positive} \mid e_{negative})$	$Pr(r_{negative} \mid e_{negative})$	*1*

Table 11.4: The probabilities of the four event-response combinations.

In addition, we can define a ratio, ***L(d)***, where ***d*** is the discriminant score of a case calculated by the discriminant function estimated from the training data set..

$$L(d) = \frac{Pr(d \mid e_{positive})}{Pr(d \mid e_{negative})}$$

L(d) summarizes the changing ratio of the corresponding ordinates of the conditional probability distributions of the discriminant scores for positive cases and negative cases.

Our purpose is to predict the sign of the effectiveness of data fusion based on the discriminant score. The rational goals of decision theory is to maximize expected value. To achieve this goal, different values (and cost, i.e., negative value) are assigned to different outcomes.

The payoff matrix is shown in *Table 11.5*, where v_1 and v_4 are positive, v_2 and v_3 are negative:

	$r_{positive}$	$r_{negative}$
$e_{positive}$	v_1	V_3
$e_{negative}$	v_2	V_4

Table 11.5: Payoff Matrix of Events and Responses

Thus, the expected value function, ***E(V)***, of a decision rule with the conditional probabilities shown in *Table 11.4* becomes:

$$\begin{aligned} E(V) = \; & Pr(r_{positive} \mid e_{positive})\, Pr(e_{positive}) \times v_1 + Pr(r_{positive} \mid e_{negative}) \times v_2 \\ & + Pr(r_{negative} \mid e_{positive})\, Pr(e_{positive}) \times v_3 \\ & + Pr(r_{negative} \mid e_{negative})\, Pr(e_{negative}) \times v_4 \end{aligned}$$

Given discriminant score ***d***, we can compute the expected value ***E(V)*** of each assignment of $r_{positive}$ and $r_{negative}$:

$$E(V \mid d, r_{positive}) = Pr(e_{positive} \mid d) \times v_1 + Pr(e_{negative} \mid d) \times v_2$$

$$E(V \mid d, r_{negative}) = Pr(e_{negative} \mid d) \times v_4 + Pr(e_{positive} \mid d) \times v_3$$

To maximize expected value, when we observe ***d***, we should predict the effectiveness of data fusion as positive if and only if

$$E(V \mid d, r_{positive}) > E(V \mid d, r_{negative})$$

that is, when

$$Pr(e_{positive} \mid d) \times v_1 + Pr(e_{negative} \mid d) \times v_2 > Pr(e_{negative} \mid d) \times v_4 + Pr(e_{positive} \mid d) \times v_3$$

If $Pr(e_{negative} \mid d)$ is not equal to 0 and v_1 is not equal to v_3, then the above inequality can be re-arranged as:

$$Pr(e_{positive} \mid d) \times (v_1 - v_3) > Pr(e_{negative} \mid d) \times (v_4 - v_2)$$

$$\Rightarrow \quad \frac{Pr(e_{positive} \mid d)}{Pr(e_{negatives} \mid d)} > \frac{v_4 - v_2}{v_1 - v_3}$$

Using Bayes' Rule, we substitute $\frac{Pr(d \mid e_{positive})Pr(e_{positive})}{Pr(d)}$ for $Pr(e_{positive} \mid d)$ and $\frac{Pr(d \mid e_{negative})Pr(e_{negative})}{Pr(d)}$ for $Pr(e_{negative} \mid d)$.

This yields:

$$\frac{Pr(d \mid e_{positive})}{Pr(d \mid e_{negative})} > \frac{Pr(e_{negative})}{Pr(e_{positive})} \times \frac{v_4 - v_2}{v_1 - v_3}$$

$$\Rightarrow \quad L(d) > \frac{Pr(e_{negative})}{Pf(e_{positive})} \times \frac{v_4 - v_2}{v_1 - v_3}$$

where ***L(d)*** is the ratio we defined above.

The above decision rule divides the ***d***-axis into two regions which may consist of several disconnected parts.. In one region we should predict the effectiveness of data fusion as positive, and in the other region we should predict the effectiveness of data fusion as negative, with the boundary point(s) point ***c*** satisfying the equation:

$$L(c) = \frac{Pr(e_{negative})}{Pr(e_{positive})} \times \frac{v_4 - v_2}{v_1 - v_3}$$

In parametric analysis we look for the best connected regions approximately this rule. In the non-parametric analysis, we permit disconnected regions. The disconnection occurs precisly when bin number 14 is added (see *Talbe 8.2*) and it is not adjacent to the bins already added to the positive region.

References

Bartell, B.T., Cottrell, G.W. & Belew R.K. (1994). Automatic combination of multiple ranked retrieval systems. Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 173-181.

Belkin, N.J., Cool, C., Croft, W.B. & Callan, J.P. (1993). The effect of multiple query representations on information retrieval performance. Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 339-346.

Egan, J.P. (1975). Signal detection theory and ROC analysis. New York: Academic Press.

Belkin, N.J., Kantor, P.B., Fox, E, and Shaw, J. (1995) Combining the evidence of multiple query representations for information retrieval. Information Processing and Management, vol. 31, No. 3, pp. 431-448.

Fox, E.A. & Shaw, J.A. (1993). Combination of multiple searches. Proceedings of the Second Text REtrieval Conference (TREC-2). National Institute of Standards and Technology Special Publication 500-215.

Fox, E.A. & Shaw, J.A. (1994). Combination of multiple searches. Proceedings of the Third Text REtrieval Conference (TREC-3). National Institute of Standards and Technology Special Publication 500-215.

Gomez, L.M., Lochbaum, C.C. & Landauer, T.K. (1990). All the right words: Finding what you want as a function of richness of indexing vocabulary. Journal of the American Society for Information Science. Vol. 41, No., 8, pp. 547-559.

Harman, D. (1993). Overview of the first Text REtrieval Conference. Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 36-48.

Harman, D. (1996). Overview of the fourth Text REtrieval Conference. In D. Harman (ed.) Proceedings of the Fourth Text Retrieval Conference. Washington. DC: GPO.

Kantor, P.B. (1994a) Data fusion in information retrieval: Towards a theoretical foundation. A vector simulation model. APLab Technical Report. APLab/TR-93/3.

Kantor, P.B. (1994b) Information retrieval technique. Annual review of information science and technology. Vol 29, pp. 53-90

Kantor, P.B. (1995) Tutorial on data fusion in information retrieval. ACM SIGIR. Seattle Washington.

Kantor P. B. (1998a) Semantic dimension: On the effectiveness of naive data fusion methods in certain learning and detection problems. APLab Technical report.

Kantor P. B. (1998b) Semantic dimension: On the effectiveness of naive data fusion methods in certain learning and detection problems. Persented at the 5th Conference on Applications of Mathematics in Artifiical Intelligence. Jane 3-4, 1998.

Kantor, P.B., Ng, K.B. & Hull, D. (In preparation) Advanced approaches to the statistical analysis of TREC information retrieval experiments. Technical report, to be published by National Institute of Standards and Technology.

Katzer, J, McGill, M.J., Tessier, J.A., Frakes, W. & Dasgupta, P. (1982). A study of the overlap among document representations. Information Technology: Research and Development. Vol. 1, No. 2, pp.261-274.

Kemeny, J.G. (1964). Random essays on mathematics, education and computers. Englewood Cliffs, N.J.: Prentice-Hall.

Lee, J.H. (1995). Combining multiple evidence from different properties of weighting schemes. Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp.180-188.

Lee, H.J. (1997). Analyses of multiple evidence combination. Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp.267-276.

Ng, K.B. and Kantor, P.B. (1996). Two experiments on retrieval with corrupted data and clean queries in TREC 4 adhoc task environment: Data fusion and pattern scanning. In D. Harman (ed.) Proceedings of the Fourth Text Retrieval Conference. Washington. DC: GPO.

Ng, K.B., Loewenstern, D., Basu, C., Hirsh, H. & Kantor, P. (1997). Data fusion of machine learning methods for the TREC-5 routing task (and other works). In D. Harman (ed.) (in press) Proceedings of the Fifth Text Retrieval Conference. Washington. DC: GPO.

Ng, K.B., Kantor, P.B. (1998). An Investigation of the Conditions for Effective Data Fusion in Information Retrieval: A Pilot Study. Proceeding of the 61th Annual Meeting of the American Society for Information Science.

Saracevic, T, & Kantor, P. (1988). A study of information seeking and retrieving. III. Searchers, searches, overlap. Journal of the American Society for Information Science. Vol. 39, No. 3, pp. 197-216.

Turtle, H. & Croft, W.B. (1991). Evaluation of an inference network-based retrieval model. ACM Transactions on Information Systems. Vol. 9, No. 3, pp. 187-222.

Varshney, P.K. (1997). Scanning the issue: Special issue on data fusion. Proceedings of the IEEE. Vol. 85, No. 1, pp. 3-5.

Viswanathan, R. & Varshney, P.K. (1997). Distributed detection with multiple sensors: Part 1 -- fundamentals. Proceedings of the IEEE. Vol. 85, No. 1, pp. 54-63.

Voorhees, E.M. & Harman, D. (1997). Overview of the Fifth Text REtrieval

Conference. In D. Harman (ed.) (in press) Proceedings of the Fifth Text Retrieval Conference. Washington. DC: GPO.

Voorhess, E.M., Gupta, N.K., & Johnson-Larid, B. (1995). Learning collection fusion strategies. In Fox, Ingwersen, & Fidel (Eds.), SIGIR'95: Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 172-179.