Middle-East Journal of Scientific Research 24 (Techniques and Algorithms in Emerging Technologies): 303-308, 2016
ISSN 1990-9233;

DOI: 10.5829/idosi.mejsr.2016.24.TAET23542

Survey on Performance Estimate Fashions for Information Retrieval System

[1]K. Seenuvasan and [2]V. Janani

[1]PG Scholar, Department of CSE, Adhiyamaan College of Engineering, Hosur, India
[2]Assistant Professor, Department of CSE, Adhiyamaan College of Engineering, Hosur, India

Abstract: The peoples are used to go for library to search any kind of information by using card catalogue method. From 20th century onwards peoples are started to search information from large stored database system. The searching of information system can also be called as information retrieval system. The efficiency of all information retrieval system has been estimate under some fashion. Today's many efficiency estimate fashions have been used. And their performance efficiency estimate fashions have been divided into two class's one is Non-Graphical Estimate Fashions and Graphical Estimate Fashions. In this paper, performance estimates for information retrieval system have been reviewed for both Non-Graphical (Precision, Recall, F1-score, MAP) and Graphical Fashions (ROC Curve, AUC and nDCG).

Key words: Precision · Recall · F1-score · MAP · ROC Curve · AUC and nDCG

INTRODUCTION

The Information retrieval (IR) system is utilized to look the data from gigantic measure of stored database. Since in 1920s the mechanical and electro mechanical based gadgets where used to inquiry data from huge measure of stored database [1]. In 1948 Holmstrom presented a "machine called the Univac" look strategy to recover the data, for example, 120 words for every minute by utilizing subject code system. The subject code incorporates the alluded data about stored information. The Holmstrom data seeking strategy made cutting edge look system, that is in 1950s the PC based data seek system was presented. From 1950 to 2000 number of PC based IR undertakings were executed, in 1960s Gerard Salton has framed IR bunch at Harvard University. This IR gathering set up thoughts and ideas about IR systems and a noteworthy accomplishment of this IR gathering was to deliver a calculation for rank based retrieval. In 1990s Berner-Lee depicted the World Wide Web, from that point the data retrieval system has confronted new sorts of issues and to take care of these issues two vital advancements were made. One is link analysis and another search of anchor text. These days, however numerous web indexes exist to recover the data by utilizing number of styles, how would we know which one is best molds? Keeping in mind the end goal to gauge IR system execution, two procedures are utilized. One is binary judgment measure [2] another is grade judgment measure [3]. A binary judgment measure is a double evaluation of relevant or non- relevant for every question record pair. The binary measures have two sorts of results one is ranked list results and unranked sets results. A grade judgment measures is an evaluation importance appraisal [3], that retrieve the relevant documents based on grade or degree of a document. The nDCG is a prominent grade judgment measures strategy that can be contemplated in future. Whatever is left of this article is sorted out as takes after. Segment 2 portrays the unranked results and Section 3 depicts the ranked results lastly, area 4 gives conclusions.

Hierarchal Structure by Performance Evaluation of Ir System: Figure 1 shows the performance estimate fashions of different IR systems. The binary judgment measure and grade judgment measure are the two procedures utilized for IR system execution assessment. The binary judgment measure creates the two sorts of assessment results one is unranked results and ranked results. Unranked results can likewise be called as Non-Graphical results. Unranked results utilize three sorts of assessment procedures that are exactness, review and F1-

Corresponding Author: K. Seenuvasan, PG Scholar, Department of CSE, Adhiyamaan College of Engineering, Hosur, India.

Scoure Ranked results support both Graphical and Non-Graphical results[4, 5] and ranked results utilize four sorts of assessment strategies that are Mean Average Precision (MAP), PR-curve, ROC-curve and AUC. The grade judgment measure delivers the ranked based results and nDCG is best suitable for assessing positioning records.

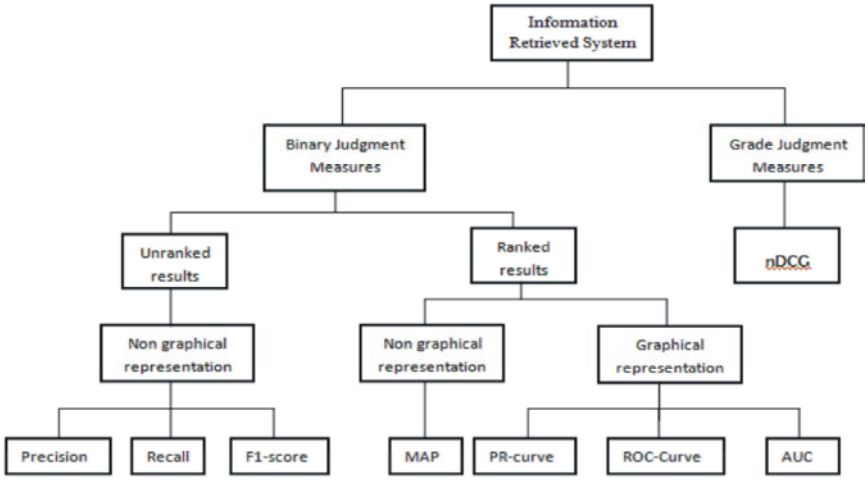

Fig. 1: Performance Estimate fashions (graphical and non graphical) hierarchal Structure.

Unranked Retrieval Results: The Unranked retrieval result is characterized as unordered records that are measured to deliver the non graphical results. The non graphical results are delivered taking into account three styles; Precision, Recall and F1-score.

Precision, Recall and F1-score: Precision is the portion of recovered reports that are relevant; recall is the part of relevant records that are recovered [6]. Precision and recall results is a parallel evaluation of either relevant (positive) or non-relevant (negative).

$$\text{Precision} = \frac{TP}{TP + FP} \quad \text{Recall} = \frac{TP}{TP+FN}$$

where,
TP-True Positive, TN-True Negative, FN-False Negative, FP-False Positive.

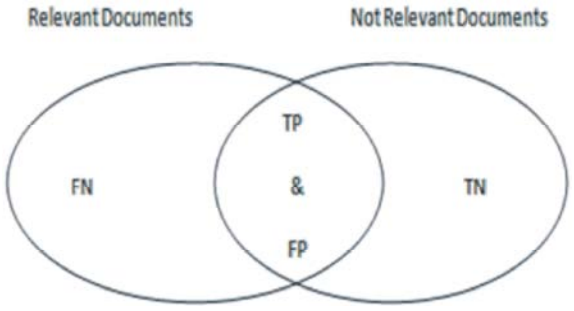

Fig. 2: Relationship between precision and recall

True positive means relevant records are recovered; True negative means non relevant records are not recovered; false positive means non relevant records are recovered; false negative means relevant records are not recovered.

precision and recall are contrarily related, when precision is high (expand), recall falls (low); when recall is high (build), precision falls (low).precision is more critical for web pursuit, recall is more essential for patent inquiry.

F-measures or F1-score is gotten from precision and recall measures. Both recall and precision joins in F1-score. The equation is given underneath,

$$F1 - score = \frac{1}{\propto * \frac{1}{P} + (1-\propto) * \frac{1}{R}}$$

where,
$\propto \in [0, 1]$ Then default means $\propto = \frac{1}{2}$

Example for Precision, Recall and F1-score:

Table 1: Example of Estimate of precision, recall and F1-score

Estimate	Relevant	Not Relevant
Retrieved	20	40
Not Retrieved	60	100

$$Precision = \frac{TP}{TP+FP} = \frac{20}{20+40} = \frac{1}{3} = 0.33$$

$$Recall = \frac{TP}{TP+FN} = \frac{20}{20+60} = \frac{1}{4} = 0.25$$

$$F1-score = \frac{1}{\alpha\frac{1}{P}+(1-\alpha)\frac{1}{R}} = \frac{2}{7} = 0.28$$

The aggregate number of records is 220; Table 1 is clarified for relevant and not relevant, recovered and not recovered records subtle elements. The precision is 0.33, recall is 0.25 and F1-Score is 0.28. Precision is ascertained by part of recovered records that are relevant; recall is computed by division of relevant records that are recovered. In view of non graphical evaluation techniques precision, recall and F1-score are utilized to deliver single scalar qualities (precision =0.33, recall =0.25 and F1-Score=0.28).

Ranked Retrieved Results: The ranked recovered result create the graphical and non graphical results, this estimation marginally reached out with unranked estimations and produce the abnormal state results. Among these outcomes precision and recall are rank based results, PR-Curve, ROC-Curve, P-Precision, MAP (Mean Average Precision), AUC.

Precision and Recall Rank Based Results: Rank based recall and precision are like unranked recall and precision, the rank based precision and recall estimation is stretched out from unranked estimations, these estimations are clarified by the given illustration.

Example 1:

Table 2: Example Data

Numbers of Documents	Name of Documents	Related Documents
1	doc1	X
2	doc123	X
3	doc456	
4	doc45	X
5	doc78	
6	doc567	X
7	doc1784	
8	doc444	
9	doc1123	
10	doc1789	X

Table 3: Values of Precision and Recall

Precision	Recall
1.00	0.20
1.00	0.40
0.66	0.40
0.75	0.60
0.60	0.60
0.60	0.80
0.57	0.80
0.50	0.80
0.40	0.80
0.50	1.00

Above Table 2 contain 10 records, there are 5 relevant records checked as "X". We recovered one by one start to finish. The primary archive is recovered; doc1 is relevant records so precision is 100% and recall are 20%. By recovering the second record, doc123 is likewise significant on the grounds that it has the same precision esteem as 100% and recall quality to 40% increments. By recovering the Third records, doc456 which is non important a precision quality is diminished to 66% and recall esteem as no progressions. When we recover the 5 relevant records recall quality has accomplished 100% yet when we measure the last accomplishment of accuracy is have just accomplished half on the grounds that 5 non relevant records are recovered.

Average Precision: The average precision worth is characterized as the average values taken by the estimations of precision. Furthermore, average precision create the single value precision and recall results.

$$\int_0^1 P(r)dr$$

where, The vital of 0 to 1 exactness is nearly approximated.

$$\sum_{k=1}^{N} P(k)\,\Delta\, r(k)$$

where, N is aggregate number of records, P (k) is the precision at a cutoff of k records, Δr (k) is the adjustment in recall that happened between cutoff k-1 and cutoff k. From Table 2 and 3 the average precision has been compute,

Average Precision
= (1.00+1.00+0.75+0.60+0.50)/5
= 0.77
Therefore the average precision value is 0.77.

Mean of Average Precision (Map): MAP is rank based Non-Graphical evaluation procedure that is utilized to deliver the non-graphical results that are significant to precision and recall. The average of the average precision esteem for an arrangement of questions is called mean average precision. Average precision is ascertained when the relevant records is recovered. The given equation clarifies the above idea.

$$MAP = \frac{1}{n(Re)} \sum_{k=1} Re_k \frac{\sum_{i=0}^{k} Re_i}{k}$$

where, n(Re) is the number of relevant records, Re_k and Re_1 take one or zero demonstrating non significant or important at position k and i separately.

Example 2:

Table 4: Example data

Numbers of Documents	Name of Documents	Related Documents
1	doc12	X
2	doc423	
3	doc45	
4	doc454	
5	doc545	
6	doc5	
7	doc725	X
8	doc445	
9	doc11	
10	doc89	X

Table 5: Values of Precision and recall

Precision	Recall
1.00	0.20
1.00	0.40
0.66	0.40
0.75	0.60
0.60	0.60
0.60	0.80
0.57	0.80
0.50	0.80
0.40	0.80
0.50	1.00

Avg. precision = (1.0+0.28+0.30)/3=0.52
Let's Consider Example 1 and 2.
The average precision of Example 1 is
Average precision = 0.77
The average precision of Example 2 is
Average precision = 0.52
MAP= (0.77+0.52)/2=0.64

The Mean Average Precision for the example 1 & 2 has been calculated and the value is 0.64.

Precision and Recall Curve: One of the great approaches to describe the execution of data retrieval system is to create the graphical path results by utilizing the precision and recall curve [6]. The given precision and recall curve chart depends on Table 2 values.

Roc Curve: Roc (Receiver Operating Characteristics) is another rank based graphical execution gauge system and ROC chart is a method for imagining, sorting out and selecting classifier in light of their execution [7, 8].

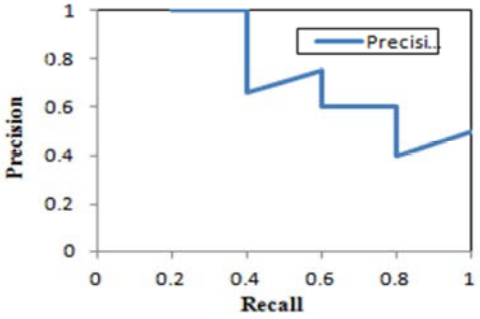

Fig. 3: The precision and recall curve for our example, It has achieved 100% recall and 50% precision

ROC curve are portrayed by two terms, one is sensitivity and specificity. Sensitivity is likewise called as recall, which is characterized as what number of relevant records have been recovered as being significant record. Specificity is what number of the not pertinent archive have been recovered as being non significant. The accompanying disarray lattice as clarified the working guideline about ROC curve.

CONFUSION MATRIX:

ROC curve confusion matrix:

		Predicated: Positive	Predicated: Negative
Ground truth	Positive	TP	FP
	Negative	FN	TN

Metrics from the confusion matrix:

Utilizing above disarray grid we need to characterized delicate and specificity took after by,

Total prediction

TP + FP = TPP, FN + TN = TPN

Total ground true TP + FN = TAP, FP+TN = TAN

True Positive Rate (recall)

$$TP\ rate = \frac{Positive\ correctly\ classified}{Total\ Positive}$$

False Positive Rate (false alarm)

$$FP\ rate = \frac{Negative\ incorrectly\ classified}{Total\ Positive}$$

$$Sensitivity = \frac{TP}{TP+FN} = \frac{TP}{TAP}$$

$$Specificity = \frac{TN}{TN+FP} = \frac{TN}{TAN}$$

Roc Space: Every single classification issue utilizes just two classes, one is positive class another is negative class, every occasion I mapped to positive "p" or negative "n" class marks. The discrete classifier model

delivers the single ROC point. Some characterization models, for example, a neural system or Naive Bayes create a constant yield. The discrete classifier model will be talked about here and remaining classification model will be examined in next area. ROC charts are two dimensional diagrams, the (TP, FP) sets are spoken to as discrete classifier. TP rate is plotted on Y pivot and FP rate is plotted on X hub [7].

We consider 100 positive and 100 negative occurrences that have been characterized in perplexity system. A, B, C ROC focuses are spoken to as discrete classifier. Figure 4 appears as every single discrete classifier. The left brings down point (0, 0) speak to as no false positive mistake furthermore no genuine positives. The right upper point (1, 1) speaks to inverse procedures of (0, 0).

'A' -Discrete classifier ROC point		
TP=63	FP=28	91
FN=37	TN=72	109
100	100	200

$$\text{TP rate} = \frac{TP}{P} = \frac{63}{100} = 0.63$$
$$\text{FP rate} = \frac{FP}{N} = \frac{28}{100} = 0.28$$

'B' -Discrete classifier ROC point		
TP=76	FP=12	88
FN=24	TN=88	112
100	100	200

$$\text{TP rate} = \frac{TP}{P} = \frac{76}{100} = 0.76$$
$$\text{FP rate} = \frac{FP}{N} = \frac{12}{100} = 0.12$$

'C' -Discrete classifier ROC point		
TP=24	FP=88	112
FN=76	TN=12	88
100	100	200

$$\text{TP rate} = \frac{TP}{P} = \frac{24}{100} = 0.24$$
$$\text{FP rate} = \frac{FP}{N} = \frac{88}{100} = 0.88$$

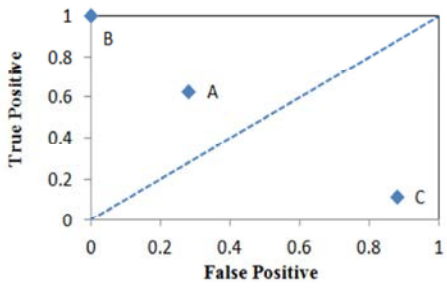

Fig. 4: A basic ROC Space (graph) showing three discrete classifiers.

The left upper point (0, 1) speaks to impeccable grouping. The "B" ROC point is immaculate execution. The right lower point (1, 0) speaks to unperfected characterization or low level exhibitions, "C" ROC point is unperfected or low execution. The most genuine spaces are overwhelmed by substantial number of negative occurrence, so execution in the far left-hand side of the ROC diagram turn out to be more interesting [8].

Create Curves in Roc Space: The discrete classifier speaks to just single point in ROC space. A few classifiers (neural system or Naive Bayes) actually yield occurrence likelihood or score [8]; the scoring classifier can be utilized with edges strategy. Every limit qualities deliver the distinctive ROC space focuses.

Table 6: Example Data for ROC

Instance	Class	Score	TP	FP
1	p	0.03	0.14	0.00
2	p	0.08	0.28	0.00
3	n	0.10	0.28	0.09
4	p	0.11	0.42	0.09
5	n	0.22	0.43	0.18
6	p	0.32	0.57	0.18
7	p	0.35	0.71	0.18
8	n	0.42	0.71	0.27
9	n	0.44	0.71	0.36
10	p	0.48	0.85	0.36
11	n	0.56	0.85	0.45
12	n	0.65	0.85	0.54
13	n	0.71	0.85	0.63
14	n	0.72	0.85	0.72
15	p	0.73	1.00	0.72
16	n	0.80	1.00	0.81
17	n	0.82	1.00	0.90
18	N	0.99	1.00	1.00

Figure 5 delineates the ROC curve of a case test set of 18 occasions, 7 positive occurrences and 11 negatives cases that are appeared in Table 6 and the examples are shorted by climbing request. The ROC focuses at (0.1, 0.7) produces its most astounding precision.

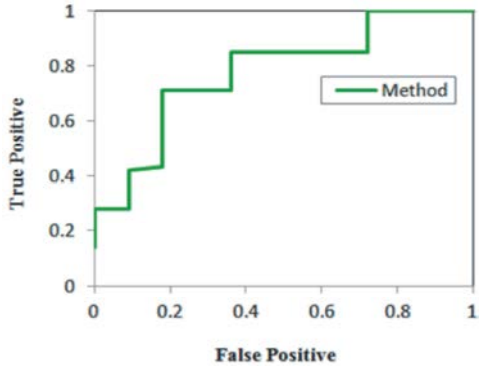

Fig. 5: Example of ROC

AUC: AUC (Area under a ROC bend) is utilized to quantify the classification models quality. The zone of the unit square is called bit of the AUC, its estimation of AUC will dependably be somewhere around 0 and 1.0. The higher AUC worth is somewhere around 0.5 and 1.0 and this worth is better quality for classification model.

CONCLUSIONS

We have exhibited a different kind of evaluation styles for data retrieval system. And have exhibited a complete binary judgment measures forms, for example, Graphical and Non-Graphical styles. The Non-Graphical designs (Precision, Recall, F1-Score, MAP) delivers the single scalar qualities. The graphical appraisal styles (PR-Curve, ROC Curve, AUC and ndcg) are utilized to picture the IR system execution for simple client view. In future work, audit on complete grade judgment measures and molds will be talked about.

REFERENCES

1. Mark sanderson and W. Bruce Croft, 2012. The history of information retrieval research", in proc. of IEEE conference, may 2012.
2. Kevin P. Murphy, 2007. Performance evaluation of binary classifiers. Technical Report, University of British Columbia.
3. Jaana Kekäläinen and Kalervo Järvelin, 2002. Using graded relevance assessments in IR evaluation, Journal of the American Society for Information Science and Technology, 53(13): 1120-1129.
4. Rasmussen, E., 2002. Evaluation in Information Retrieval, in 3rd International Conference on Music Information Retrieval, Paris, France, pp: 45-49.
5. Zuva, K. and T. Zuva, 2012. Evaluation of Information Retrieval Systems, International Journal of Computer Science & Information Technology (IJCSIT), 4: 35-43.
6. Davis, J. and M. Goadrich, XXXX. The Relationship between Precision-Recall and ROC Curves, Proc. Int?l Conf. Machine Learning, pp: 233-240.
7. Fawcett, T., 2006. An Introduction to ROC Analysis, Pattern Recognition Letters, 27(8): 861-874.
8. Fawcett, T., 2003. ROC Graphs: Notes and Practical Considerations for Data Mining Researchers, 2003 :HP Labs