An Analysis of Ranking Principles and Retrieval Strategies

G. Zuccon* and L. Azzopardi and C.J. van Rijsbergen
{guido, leif, keith}@dcs.gla.ac.uk

School of Computing Science
University of Glasgow
Scotland, UK

Abstract The assumptions underlying the Probability Ranking Principle (PRP) have led to a number of alternative approaches that cater or compensate for the PRP's limitations. All alternatives deviate from the PRP by incorporating dependencies. This results in a re-ranking that promotes or demotes documents depending upon their relationship with the documents that have been already ranked. In this paper, we compare and contrast the behaviour of state-of-the-art ranking strategies and principles. To do so, we tease out analytical relationships between the ranking approaches and we investigate the document kinematics to visualise the effects of the different approaches on document ranking.

1 Introduction

The Probability Ranking Principle (PRP) has played a central role in the development of Information Retrieval (IR). The PRP has largely stood the test of time for adhoc retrieval, but for emerging retrieval tasks, such as novelty and diversity, the assumptions made by the PRP have been shown to lead to non-optimal performance [2,5,7]. Alternative ranking approaches have been proposed; these include two ranking strategies, Maximal Marginal Relevance (MMR) [1] and Portfolio Theory (PT) [7], along with the Quantum PRP (qPRP) [8], and the Interactive PRP (iPRP) [3]. Each approach can be regarded as a revision of the PRP, where the point of departure is the introduction of document dependent evidence within the revised ranking. The function used for revising a ranking may be formulated differently, depending upon the ranking approach. However, the net effect of the revision boils down to the promotion of diversity, i.e. documents which are different from those previously seen in the ranking are promoted up in the ranking, or of similarity, i.e. documents that are similar to the previous one, obtaining a sort of pseudo-relevance feedback effect.

While there has been a lot of interest in this area and a number of empirical comparisons, there has been no formal analysis of these approaches. Given that these new approaches attempt to address the same problem, it is important to identify specifically and formally relationships, similarities and differences

* Supported by EPSRC Grant number EP/F014384/ and Zirak s.r.l. (http://www.zirak.it/)

between methods, in order to contextualise existing methods and to develop improved theory.

To this end, we perform a comprehensive theoretical analysis and comparison of ranking principles and strategies. We first introduce each approach in section 2, establishing a common framework, which allows us to further contrast them from an analytical perspective. Indeed, in section 3 we tease out relationships among approaches by analysing their ranking behaviour within a small scale controlled scenario. The analysis is completed in section 4 where we investigate the document kinematics that different approaches impose on the rankings.

2 Principles and Strategies

Approaches to ranking can be divided into two categories:

strategies that are empirically driven and devised to cater for the limitations of the PRP, i.e. Maximal Marginal Relevance [1] and Portfolio Theory [7], and,

principles that are theoretically driven and implicitly cater for the limitations of the PRP, i.e. the interactive PRP [3] and quantum PRP [8].

Regardless of the approach, strategy or principle, the recently proposed alternatives to the PRP mathematically deviate through the inclusion of a function that captures dependencies between documents. This function expresses the relationship between documents: depending upon how the function is set, the ranking approach promotes either document diversity or similarity. As we shall see, alternatives differ in the way dependencies are incorporated, and the extent of parameterisation of the ranking formula. Specifically, PT and qPRP are characterised by an additive ranking function, MMR by an interpolated and iPRP by a multiplicative, where PT and MMR are by definition parameterised. On the contrary, in their original formulations iPRP and qPRP do not have parameters. However, parametric instantiations may be formulated as well for qPRP and iPRP.

Next we will provide the formal analysis to justify the previous statements by providing a common framework to describe each of the principles and strategies, so that we can compare them analytically in a straightforward manner.

Probability Ranking Principle

The PRP states that documents should be retrieved in decreasing order of their estimated probability of relevance given the query [6]. By adhering to the PRP, at each rank position i the IR system should select a document d_i such that:

$$d_i = \underset{d \in \mathcal{RE} \setminus RA}{\arg\max}\ P(d) \tag{1}$$

where $P(d)$ is the probability of a document being relevant given the query, RA is the list of documents that have been ranked, and d is a document belonging to the set of retrieved documents ($\mathcal{RE}$). Ranking according to this criteria has been shown to provide the optimal ranking [6]. This, however, depends upon a number of assumptions; of those the most criticised are:

(i) the independent assessment of document relevance (i.e. independence assumption); and

(ii) the certainty of the estimation of relevance.

Goffman noticed that by assuming independence between document's relevance assessments, the "relationship between query and the document is both necessary and sufficient to establish relevance" [4]. It has been argued [2,5] that this is not strictly the case in real search scenarios, where document's relevance depends upon information acquired during the course of the retrieval process. Goffman formalised this intuition as follows: the relevance of a document must depend upon what is already known at the time the document is examined by the user. If a document d has been judged relevant to a particular information need, the relevance of other documents might be affected by the relevant information already known. Gordon and Lenk have demonstrated the sub-optimality of the PRP when the independence assumption does not hold [5]. While, Chen and Karger showed that the PRP is not always optimal for different information needs [2]. These limitations and a number of empirical observations regarding the PRP have motivated a number of alternative ranking strategies and principles.

2.1 Alternatives to the PRP

In the following we consider ranking approaches alternative to the PRP. A common trend between these alternatives is the presence in the ranking function of two main elements: (1) the probability of relevance, or score of the document; and (2) a function that estimates the similarity between the representations of two documents. To facilitate comparison, we reformulate the approaches in a common framework, so that their ranking formulas are written with respect to a common estimation of the probability of relevance for a document d (represented by $P(d)$), and a common similarity function between documents. In the following we select the Pearson's correlation coefficient[1] $\rho_{d,d'}$ as measure of similarity between d and d'.

Maximal Marginal Relevance

In Maximal Marginal Relevance (MMR) [1], an hyper-parameter λ is used to balance the similarity between document and query, and the similarity between the candidate document and documents ranked at earlier positions. A document at rank i is selected using the following objective function:

$$d_i = \underset{d \in \mathcal{RE} \setminus RA}{\arg\max} \left(\lambda s(d,q) - (1-\lambda) \max_{d' \in RA} sim(d,d') \right)$$

where $s(d,q)$ is a similarity function between document and query, while $sim(d,d')$ is a function that determines the similarity between documents d and d'. If two candidate documents have the same probability of relevance (or $s(d,q)$), MMR will rank first the one that is least similar to any of the documents that have been ranked at previous positions. The hyper-parameter can be inferred by the user's model: $\lambda < 0.5$ characterises users with a preference for rankings where document dependencies are more important than relevance. Greater values of λ

[1] This choice is motivated by the fact that Pearson's correlation is used within PT and in previous instantiations of the qPRP. The choice of similarity function across all ranking approaches is however rather arbitrary: we have kept them all the same so that the quintessential differences between approaches can be teased out.

would capture the converse situation. For consistency, we re-state MMR in terms of $P(d)$ and $\rho_{d,d'}$ in place of $s(d,q)$ and $sim(d,d')$, respectively:

$$\textbf{MMR: } d_i = \arg\max_{d \in \mathcal{RE} \setminus RA} \left(\lambda P(d) - (1-\lambda) \max_{d' \in RA} \rho_{d,d'} \right) \tag{2}$$

Portfolio Theory

Portfolio Theory applied to IR [7] attempts to minimise the risk associated with ranking documents under uncertainty in their relevance estimates by balancing the expected relevance value (mean) and its variance. The ranking criteria combines the estimated document relevance with (i) an additive term which synthesises the risk inclination of the user, (ii) the uncertainty (variance) associated with the probability estimation, and (iii) the sum of the correlations between the candidate document and documents ranked in previous positions. For each rank position i, documents are selected according to:

$$\textbf{PT: } d_i = \arg\max_{d \in \mathcal{RE} \setminus RA} \left(P(d) - b w_d \sigma_d^2 - 2b \sum_{d' \in RA} w_{d'} \sigma_d \sigma_{d'} \rho_{d,d'} \right) \tag{3}$$

where b encodes the risk propensity of the user, σ_d^2 is the variance associated to $P(d)$, and w_d is a weight that expresses the importance of the rank position of d and d'. When PT has been employed in practice, σ_d^2 has been treated as a model parameter (see [7,8]), because a single point-wise relevance estimation is used: in the rest of the paper we follow the same route.

Interactive PRP

In [3], Fuhr proposes a theoretical framework for extending the PRP to the context of interactive IR where the independence assumption is rejected. This is because in interactive searches relevance depends on documents the user has previously examined. Search is therefore modelled as situation, i.e. a list of choices the user is presented with: users move between situations by accepting one of the choices they are provided with. Once a choice is accepted, the retrieval system produces a new list of choices dependent from the previous choice. The ranking principle strives to provide the optimal ordering of the choices presented in each situation. For each rank i, documents under the iPRP are ranked as follows:

$$d_i = \arg\max_{d \in \mathcal{RE} \setminus RA} \left[e + P(d) \left(b_{d,i} Q(d) + g(1 - Q(d)) \right) \right] \text{ , where}$$

- $Q(.)$ is the probability that the user does not revise their choice of selecting document d (i.e. the probability that the user does not change their mind about the relevance of the document d after examining it);
- e is the effort of examining document d;
- g is the additional effort required for correction if the user judges a viewed document as irrelevant;
- $b_{d,i}$ is the benefit of ranking document d at rank i if the document is relevant.

In this study, we provide a possible instantiation of the iPRP for the first pass of retrieval (i.e. before any actual user interaction has transpired): in this context we do *not* consider any further interaction or re-ranking. This instantiation is in line with the assumptions of [3], and had been first proposed in [10]. Since we are examining the case of the first pass of retrieval, we assume e, g and $Q(.)$ as constants. These can then be dropped for rank equivalence reasons. We then consider the benefit of ranking d at rank i. A reasonable approximation would be to determine how similar the current candidate document is with all previous documents. This is because $b_{d,i}$ is dependent upon previously ranked documents. We achieve this through a summation over all previously ranked documents of the negative correlation[2] between previously ranked documents and d. If document d is similar to previous documents, then the correlation will be low, and possibly negative: the total benefit achieved will thus be low. Similar documents are demoted in the ranking, while diverse documents are promoted, giving rise to the following objective function:

$$\textbf{iPRP:}\ d_i = \operatorname*{arg\,max}_{d \in \mathcal{RE} \setminus RA} \left(P(d) b_{d,i}\right) = \operatorname*{arg\,max}_{d \in \mathcal{RE} \setminus RA} \left(-P(d) \frac{\sum_{d' \in RA} \rho_{d,d'}}{|RA|}\right) \tag{4}$$

Under the iPRP dependencies between documents are incorporated through *multiplication*, providing a completely different approach to the other alternatives.

Quantum PRP

The qPRP develops from quantum probability theory (as opposed to traditional Kolmogorovian probability theory), and naturally incorporates dependencies between documents through the notion of *quantum interference* [8]. In order to obtain the most valuable document ranking for a user the total probability of relevance of the ranking needs to be maximised. The interference $I_{d,d'}$ between two documents influences the total probability of relevance (see [8]). The qPRP then selects a document d to be ranked at position i such that:

$$d_i = \operatorname*{arg\,max}_{d \in \mathcal{RE} \setminus RA} \left(P(d) + \sum_{d' \in RA} I_{d,d'}\right)$$

The underlying intuition is that documents in a ranking share relationships at relevance level, i.e. they interfere with each other, and the interference has to be taken into account when ranking documents. According to [8], interference can be approximated via a function such as the correlation $\rho_{d,d'}$ between documents[3], where $I_{d,d'} = -2\sqrt{P(d)}\sqrt{P(d')}\rho_{d,d'}$. Therefore, the ranking rule becomes:

$$\textbf{qPRP:}\ d_i = \operatorname*{arg\,max}_{d \in \mathcal{RE} \setminus RA} \left(P(d) - 2 \sum_{d' \in RA} \sqrt{P(d)}\sqrt{P(d')}\rho_{d,d'}\right) \tag{5}$$

2.2 Parametric Instantiations of iPRP and qPRP

While MMR and PT are by definition characterised by the settings of their parameters, the instantiations of iPRP and qPRP of Eqs 4 and 5 are not parametric. However, parametric instantiations of these principles can be given, where parameters control the impact of correlation on the ranking process. The parameter is formally introduced within the approximations of benefit and interference.

---

[2] A negative value implies a cost to the user. This might occur when examining relevant but redundant information.

[3] While $\sqrt{P(d)}\sqrt{P(d')}$ is the magnitude of the complex probability amplitudes associated to documents d and d'.

When instantiating the iPRP, the benefit of ranking a document d at rank i (i.e. $b_{d,i}$) has been approximated as $-\frac{\sum_{d' \in RA} \rho_{d,d'}}{|RA|}$. A possible parametric instantiation of the iPRP is obtainable by setting $b_{d,i} = -\beta \frac{\sum_{d' \in RA} \rho_{d,d'}}{|RA|}$, with β being a free parameter (and $\beta \in \mathbb{R}$). Therefore, the ranking formula of iPRP becomes:

$$\textbf{iPRP(parametric):}\ d_i = \arg\max_{d \in \mathcal{RE} \setminus RA} \left(-\beta P(d) \frac{\sum_{d' \in RA} \rho_{d,d'}}{|RA|} \right) \tag{6}$$

Similarly, when operationalising the qPRP, interferences have been approximated as $I_{d,d'} = -2\sqrt{P(d)}\sqrt{P(d')}\rho_{d,d'}$. Alternative approximations have been investigated in [9]: these considered similarity functions other than Pearson's correlation for estimating interferences and no parameter was introduced. We can however consider a parametric instantiation of the qPRP, by introducing the parameter β in the approximation of the interference term, obtaining:

$$\textbf{qPRP(parametric):}\ d_i = \arg\max_{d \in \mathcal{RE} \setminus RA} \left(P(d) - 2\beta \sum_{d' \in RA} \sqrt{P(d)}\sqrt{P(d')}\rho_{d,d'} \right) \tag{7}$$

The first contribution of this paper is the common framework for describing ranking approaches. Using this framework we can now perform an analysis of their ranking behaviour and of the kinematics imposed on relevant documents.

3 Analysis of Ranking Behaviours

Each approach handles document dependencies in a characteristically different way. The question is: *How do different approaches affect document ranking?*

To answer this question, we shall consider two aspects: (1) what document is ranked first?, and (2) what documents are then subsequently ranked next?

For all approaches, the document ranked at fist position (i.e. $i = 1$) is the same. This is the document which has the highest probability of relevance. Differences between alternatives and the PRP manifest at ranks greater than one. At $i > 1$, each alternative approach will tend to revise the original ranking such that documents which are different to those ranked previously will be promoted. To obtained deeper intuition of this phenomena for each ranking alternative, we analytically compare each method at the functional level to determine more precisely how the ranking of documents would be affected.

To this aim, we shall consider the following example scenario, where we have two documents, d and d', with the same probability of relevance, i.e. $P(d) = P(d')$, and d has been ranked first. We are interested to determine what is likely to happen to d' given the PRP, MMR, PT, iPRP, and qPRP: i.e. is it likely to be demoted or promoted? We consider three further cases, where documents d, d' are:

case 1: virtually identical[4] and thus positively correlated, i.e. $\rho_{d,d'} = 1$;

[4] We consider the document term vectors to compute correlations (and thus dependencies): term-position does not influence correlation, while term's (weighted) presence does. Two documents containing the same exact text, but shuffled in different orders, will appear identical to the correlation function.

case 2: with nothing in common, and thus not correlated at all, i.e. $\rho_{d,d'} = 0$;

case 3: sharing the same terms, but with complete different use and frequencies, and thus anti-correlated[5], i.e. $\rho_{d,d'} = -1$.

Probability Ranking Principle The behaviour of the PRP does not depend on the correlation, so the PRP always ranks documents d and d' consecutively, and actually both $(d, d', ...)$ and $(d', d, ...)$ are valid rankings.

Maximal Marginal Relevance When documents are correlated (case 1), MMR assigns to d' the score $\lambda P(d') - (1 - \lambda)$, which might assume negative values. If $\lambda = 1$ then MMR reduces to PRP, while if $\lambda = 0$ document d' gets a score of 1. For $0 < \lambda < 1$, the original score of $P(d')$ is remodulated by λ and then decreased of $(1 - \lambda)$. In case 2, MMR rescales the document's probability by the hyper-parameter, assigning to d' the score $\lambda P(d')$. The document score increases in the third case, i.e. when the correlation has negative value, adding to the (re-scaled) probability of the document a value proportional to $1 - \lambda$: if $\rho_{d,d'} = -1$, then the score of d' is $\lambda P(d') + 1 - \lambda$.

Portfolio Theory The score PT assigns to a document differs to the one provided by the PRP of $-bw_d\sigma_d^2 - 2bw_{d'}\sigma_d\sigma_{d'}\rho_{d,d'}$. The sign of PT's variation in scores, i.e. increment or decrement, are then not only dependent upon the correlation's sign, but also upon the user's model parameter b. We focus our analysis on the situation where $b > 0$: under this circumstance PT promotes diversity in the document ranking. The initial document probability of relevance is revised of $-|b|w_d\sigma_d^2 - 2|b|w_{d'}\sigma_d\sigma_{d'}\rho_{d,d'}$. In case 1, i.e. $\rho_{d,d'} = 1$, the score of d' is decreased by $-|b|w_d\sigma_d^2 - 2|b|w_{d'}\sigma_d\sigma_{d'}$. If documents are not correlated (case 2), the initial score undergoes a limited decrement of $|b|w_d\sigma_d^2$. Finally, in case 3 (anti-correlated documents), the initial score of d' is modified by PTs's ranking formula of $-|b|w_d\sigma_d^2 + 2|b|w_{d'}\sigma_d\sigma_{d'} \approx |b|\sigma_d^2(2w_{d'} - w_d)$. The discount factor w_d is estimated through a monotonically decreasing function of the document's rank position, thus $2w_{d'} - w_d$ can be either positive or negative. If positive, d''s score gets incremented; vice versa, d' gets demoted in the document ranking. Finally, when $b = 0$ PT's ranking function reduces to the one of the PRP.

Interactive PRP The iPRP is characterised by a multiplicative ranking function. When d and d' are completely correlated (case 1), iPRPs assigns to d' the score $-P(d')$, and thus the document is demoted: documents that are more relevant than others would suffer a stronger demotion. In the situation of zero-correlated documents (case 2), d' gets assigned a score of zero and is demoted in the ranking. In case 3, iPRPs assigns to d' the same score obtained with the PRP, i.e. $P(d')$, and thus d' is ranked immediately after d (as in the PRP).

Quantum PRP When documents correlate, as in case 1, the probability assigned to d' is revised and is modified to the value $-P(d')$: this is due to the interference term becoming $I_{d,d'} = -2\sqrt{P(d)}\sqrt{P(d')} = -2P(d')$. In this situation, as for other models, also according to the qPRP d''s chances to get

[5] While in practice correlations of -1 are unlikely, there might be cases where correlations are negative because of the weighting schema used to compute document term vectors. However, for the purpose of our example, we imagine the two documents to be completely anti-correlated.

Table 1. Overview of the characteristics of the ranking principles and strategies.

Model	Dependence	Parameters	$\rho = 1$	$\rho = 0$	$\rho = -1$
PRP	-	-	$\circ$	$\circ$	$\circ$
MMR	Interpolated	λ: hyperparameter	$\downarrow$	$\sim$PRP	$\uparrow$
PT	Additive	b: user risk propensity σ: variance estimation relevance w: discount rank position	$\downarrow$ (if $b > 0$)	$\sim$PRP	$\uparrow$ (if $b > 0$)
iPRP	Multiplicative	-	$\downarrow$	0	$\uparrow$
qPRP	Additive	-	$\downarrow$	=PRP	$\uparrow$

ranked at second position are decreased, possibly demoting it to lower positions. When d and d' are not correlated at all as in case 2, i.e. $\rho_{d,d'} = 0$, qPRP does not change PRP's estimate since the interference term is zero: there is no dependence between the actual candidate and the previous ranked document. In case 3, qPRP boost the original probability of d' to the quantity $3 \cdot P(d')$. In fact, the interference term results $I_{d,d'} = 2\sqrt{P(d)}\sqrt{P(d')} = 2P(d')$.

Summary The approaches revealed a common pattern. When promoting diversity, the initial probability estimation associated to d', i.e. $P(d')$, is revised by a quantity proportional to the correlation of d' with those documents that have been already ranked. The revision increments the initial probability estimation if documents are anti-correlated. Vice versa if documents are correlated, the document score is decreased. The case of no correlation (case 2) is handled differently by each ranking approach: for example iPRP assigns to the document a zero score, while qPRP returns the same probability estimation of PRP.

Finally, the amount of revision that the score of a document is subject to depends upon the parametrisation of the ranking function. Specifically:

- MMR weights the contribution of the correlation depending on λ; high values of λ (i.e. $\lambda \to 1$) return rankings similar to those of PRP;
- PT modulates the contribution of the correlation by the product of the parameters b and σ_d^2, and considering the importance of the rank position;
- iPRP reduces the influence of the correlation by a quantity inversely proportional to the number of documents retrieved at previous ranks;
- qPRP modulates the contribution of the correlation by the square root of the probabilities of the documents involved in the comparison.

4 Kinematics of Documents

To provide a deeper understanding of the revision process, in the following we empirically explore the movement of the relevant documents.

To do so, we employ the Clueweb09 collection (part B only) and the TREC 2009-2010 Web Diversity topics and relevance judgements. Documents and queries were stemmed and stop-words were removed: thereafter documents were indexed using the Lemur 4.10 toolkit[6]. Documents were retrieved according to a unigram language model with Dirichlet smoothing ($\mu = 2,500$): for each query, the 100 documents with higher score were considered for ranking. The PRP

[6] http://lemurproject.org/

ranking was formed arranging documents in decrease order of scores. Approaches alternative to the PRP were used to re-rank documents. For PT, we regarded both the variance of the probability estimations (σ^2) and b as parameters, and we let them varying in the ranges $[10^{-7}, 10^{-2}]$ (with decimal increments) and $[-10, +10]$ (with unitary increments), respectively. MMR's hyper-parameter was varied in the range $[0, 1]$ with steps of 0.1. We considered the parametric versions of iPRP and qPRP (Eqs. 6 and 7), studying values of β varying in the range $[-1, 1]$ with steps of 0.1. Pearson's correlation between (normalised) term frequency representations of documents was employed in all re-ranking approaches.

For each ranking approach, we built a retrieval run by tuning the parameters with respect to α-NDCG@10[7] on a query-by-query basis: that is, for each query, we rank documents using the best parameter values for the query.

While our focus is on the kinematics of documents, we report the performance of the runs, to show how the re-ranking affects performance. Specifically, the approaches obtained the following values of α-NDCG@10 [8]:

$$\text{PRP: } 0.137 < \text{qPRP: } 0.172^* < \text{PT: } 0.182^* < \text{iPRP: } 0.197^* < \text{MMR: } 0.205^*$$

To illuminate the differences in the re-ranking strategies, we focus on the kinematics of only the relevant documents. In particular, for each ranking approach, we recorded the change in the position of each relevant[9] document between the alternative ranking approach and the PRP. We thus count the number of times and the extent of the promotion or demotion of relevant documents with respect to the PRP. In Figure 1 we plot the distributions of the (relevant) document kinematics, where on the x-axis zero indicates no movement of documents, greater than zero indicates that the documents have been promoted, while lesser than zero indicates the documents have been demoted. The y-axis shows the frequency of the movement. To assess the symmetry of the kinematics shapes with respect to the zero-movement abscissa (i.e. the zero on the x-axis) we consider the area under the curve (AUC), that is given by the sum of the frequencies of promotions or demotions for a given approach. Specifically, we define as AUC left (AUCL) the sum of the frequencies for $x \in [-100, -1]$, while AUC right (AUCR) is defined as the sum of the frequencies for $x \in [+1, +100]$. We further extend the notion of AUC to a weighted version (WAUC) which weights each movement amplitude (each x value on the x-axis) by its frequency $f(x)$ and normalises this by the number of movements amplitudes different from zero contained in the considered movement range (note that for some values of x there is no movement). Formally, WAUC for a range $\mathcal{R}$ is defined as:

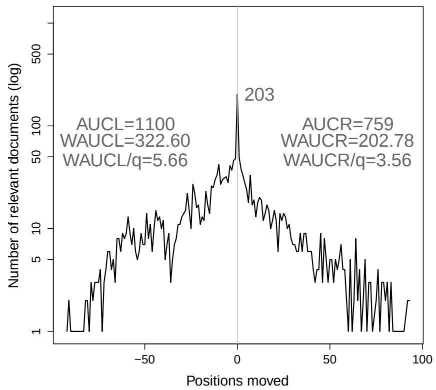
$$WAUC(R) = \frac{\sum_{x \in \mathcal{R}} |f(x) \cdot x|}{\sum_{x \in \mathcal{R}} \upsilon(x)} \text{ , where } \upsilon(x) = \begin{cases} 1 & \text{if } f(x) > 0 \\ 0 & \text{otherwise} \end{cases}$$

[7] With $\alpha = 0.5$, set according to the TREC 2009 and 2010 Web Track guidelines.

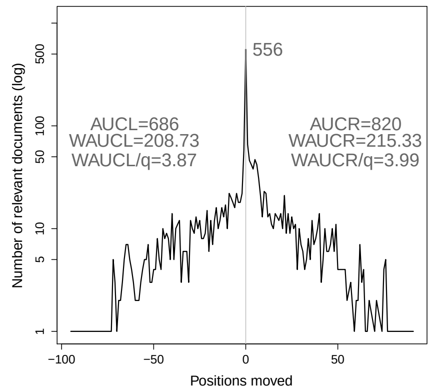
[8] Where * indicates statistical significant differences with respect to the PRP as measured by a two tailed paired t-test with $p \ll 0.01$. Note that no statistical significant differences were found between the performances of PT, MMR, iPRP and qPRP.

[9] We considered a document relevant if it is relevant to at least one facet/intent.

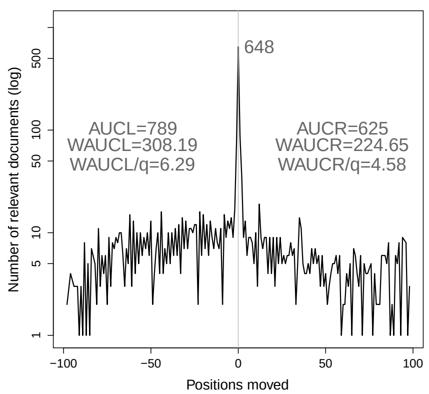
Figure 1. Kinematics, with respect to the PRP, imposed to the relevant documents by ranking strategies that cater for document dependencies. We also report the values of AUC, WAUC and the WAUC-to-query ratio (WAUC/q). Finally, in correspondence to $x = 0$, we report the frequency of zero-movements, i.e. $f(x = 0)$.



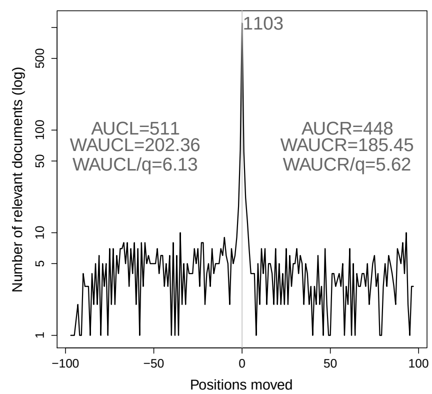
(a) Kinematics for MMR.



(b) Kinematics for PT.



(c) Kinematics for iPRP.



(d) Kinematics for qPRP.

In particular, in the following we consider WAUCL for $x \in [-100, -1]$ (the area on the left of the zero-movement abscissa) and WAUCR for $x \in [+1, +100]$ (the area on the right of the zero-movement abscissa). Values of (W)AUCL and (W)AUCR for each approach are reported in Figure 1, together with the frequency of the zero-movement (i.e. $f(x = 0)$).

Retrieval strategies (i.e. PT and MMR, Figures 2(a) and 2(b)) are characterised by wider kinematics shapes that the ones of the principles (i.e. iPRP and qPRP, Figures 2(c) and 2(d)). MMR appears to be the approach that most revises the position of relevant documents, as it is characterised by the lowest frequency of zero-movements among all approaches. This might be mainly due to the fact that for 57 out of the 98 queries of the TREC 2009-2010 dataset the best performing value of the parameter λ is different from 1: that is, MMR's ranking function effectively provides a ranking different than that of PRP, while for the remaining 41 queries MMR's ranking function reduces to PRP's one (since $\lambda = 1$

for these queries). The movement of relevant documents that is witnessed in Figure 2(a) is therefore generated by a high number of queries. While, movements that form the kinematics shapes of other approaches involve a lower number of queries. Specifically, the number of queries for which the best performing parameters do not reduce the ranking functions to that of PRP are 54 for PT, 49 for iPRP, 33 for qPRP.

The shape of MMR's kinematics is asymmetric and unbalanced towards the left side of the x-axis. The AUC of MMR confirms this impression: AUCL amounts to 1100, while the AUCR amounts to 759. This suggests that relevant documents are demoted more times than what are promoted. If compared to the kinematics shapes of other approaches, that of MMR can be regarded as being the most unbalanced towards the left side of the x-axis. Nevertheless, MMR achieves the highest value of α-NDCG@10 in our experiments: this might be because the relevant documents that are most demoted are those that are also most redundant, while the relevant documents that get promoted are novel with respect to the ones ranked at previous positions.

The shape of PT's kinematics is similar to the one of MMR's, although PT moves less relevant documents than MMR (higher zero-movement frequency) and its kinematics "ends" sooner than MMR's: no relevant documents are moved of more than 90 positions up or down the ranking. Furthermore, the kinematics of PT seems to favour the promotion of relevant documents over their demotion, as the kinematics shape is slightly unbalanced towards the right of the x-axis. This is confirmed by the difference between AUCR and AUCL; note that PT is the only approach for which AUCR $>$ AUCL. However, the difference between the area under the curve for the left and the right range decreases if WAUC is considered (i.e. WAUCL= 208.73, WAUCR= 215.33): this means that PT promotes relevant documents of fewer positions more than the ones it demotes.

The kinematics of the ranking principles (i.e. iPRP and qPRP) have a common shape. The kinematics are characterised by a high spike in correspondence of the zero-movement coordinate and a fast flattering out shape when movements involve more than half a dozen rank positions (note that the y-axis is in log-scale). The central spike represents no movement of relevant documents with respect to PRP: more relevant documents are moved by iPRP than qPRP. As for MMR, this observation is in line with the number of queries for which iPRP and qPRP provide a ranking different than PRP's one: this happens 49 times (out of 98 queries – i.e. for the 50% of the cases) for iPRP, while only 33 times for qPRP. For both principles the shapes are asymmetric and slightly unbalanced towards left (AUCL $>$ AUCR).

By comparing the WAUC of the approaches' kinematics, we can understand which strategy promotes or demotes relevant documents of more positions. Note however that a higher WAUC might not be due only to a propensity to promote or demote relevant documents of more positions, but might be as well biased by the number of queries that generated the kinematics. A better indication might be provided by the WAUC-to-query ratio (reported in Figure 1), where WAUC is divided by the number of queries for which there has been an effective movement of relevant documents with respect to the PRP. For example, while WAUCR of PT (215.33) is higher than the one of qPRP (185.45), WAUCR-to-query ratio of PT (3.99) is lower than the correspondent value for qPRP (5.62).

Notably, the lowest WAUC-to-query ratio is achieved by MMR with respect to documents that are promoted up the ranking (see WAUCR/q ratio of MMR), suggesting that overall MMR is the approach that less promotes relevant documents. However, MMR is not the approach that most demotes relevant documents, as the WAUCL-to-query ratios of iPRP (6.29) and qPRP (6.13) are higher than that of MMR. The highest promotion of relevant documents is achieved by qPRP (WAUCR/q = 5.62): however this positive characteristic does not seem to find a parallel in the retrieval performances (at least in terms of α-NDCG@10). This might be due to the fact that (i) promoted relevant documents are redundant with respect to those ranked at previous positions, and/or (ii) promotions of relevant documents do not take place within the first 10 rank positions.

The previous analysis clearly shows how each ranking approach moves relevant documents within the ranking. As a further note, we can observe that if little movement transpires then the retrieval results are similar to the PRP, while more movement results in greater or lower performance.

5 Summary and Future Work

In this paper, we have investigated a number of ranking strategies and principles that have been proposed in the literature. Our analysis focused both on the analytical relationships between the approaches and on their ranking behaviours. We have shown the links that exist between ranking approaches. Moreover we have described the behaviours of the approaches when having to decide whether promote or demote a document given previously ranked evidence. Finally, we have examined the relevant document kinematics with respect to the PRP that the re-ranking approaches impose on the ranking: to the best of our knowledge, this is the first work that investigates this aspect of ranking approaches.

References

1. J. Carbonell and J. Goldstein. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *SIGIR '98*, 335–336, 1998.
2. H. Chen and D. R. Karger. Less is more: probabilistic models for retrieving fewer relevant documents. In *SIGIR '06*, 429–436, 2006.
3. N. Fuhr. A probability ranking principle for iir. *JIR*, 12(3):251–265, June 2008.
4. W. Goffman. On relevance as a measure. *Info. Stor. & Ret.*, 2(3):201–203, 1964.
5. M. D. Gordon and P. Lenk. When is the prp suboptimal. *JASIS*, 43(1):1–14, 1999.
6. S. E. Robertson. The probability ranking principle in IR. *J. Doc.*, 33:294–304, 1977.
7. J. Wang and J. Zhu. Portfolio theory of information retrieval. In *SIGIR '09*, 115–122, 2009.
8. G. Zuccon and L. Azzopardi. Using the quantum probability ranking principle to rank interdependent documents. In *ECIR 2010*, 357–369, 2010.
9. G. Zuccon, L. Azzopardi, C. Hauff, and C.J. van Rijsbergen. Estimating interference in the QPRP for subtopic retrieval. In *SIGIR '10*, 741–742, 2010.
10. Guido Zuccon, Leif Azzopardi, and C.J. van Rijsbergen. The interactive PRP for diversifying document rankings. In *SIGIR '11 (to appear)*, 2011.