Back to the Roots: Mean-Variance Analysis of Relevance Estimations

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

School of Computing Science, University of Glasgow

Abstract Recently, mean-variance analysis has been proposed as a novel paradigm to model document ranking in Information Retrieval. The main merit of this approach is that it diversifies the ranking of retrieved documents. In its original formulation, the strategy considers both the mean of relevance estimates of retrieved documents and their variance. However, when this strategy has been empirically instantiated, the concepts of mean and variance are discarded in favour of a point-wise estimation of relevance (to replace the mean) and of a parameter to be tuned or, alternatively, a quantity dependent upon the document length (to replace the variance). In this paper we revisit this ranking strategy by going back to its roots: mean and variance. For each retrieved document, we infer a relevance distribution from a series of point-wise relevance estimations provided by a number of different systems. This is used to compute the mean and the variance of document relevance estimates. On the TREC Clueweb collection, we show that this approach improves the retrieval performances. This development could lead to new strategies to address the fusion of relevance estimates provided by different systems.

1 Introduction

I recent works, mean-variance analysis has been proposed to address the problem of document ranking in Information Retrieval [1,2,3,4]. This proposal led to the introduction of a new document ranking strategy, known as Portfolio Theory for Information Retrieval, that aims to balance the mean of estimations of (probability of) document relevance and their variance. However, while the theoretical model considers that a number of relevance estimations are available for each document and it exploits variances to revise a document ranking; its empirical instantiations refrain to use the actual mean and variance of such estimations (e.g. [1,2]). This is because the function used to estimate document relevance does not provide a probability distribution, but rather assigns to each document a single point-wise estimation of that probability. In the latter situation, there is no sense in computing a mean or a variance. The reason being, the mean of a single point-wise estimation is just equal to the estimation itself, while the variance is null. Thus, in previous works, mean and variance are discarded. In particular, the mean is replaced by the probability estimation of relevance itself. While, the variance is treated either as one of the parameters of the model, and thus needs to be estimated, or as a quantity dependent on document length. Furthermore, the ranking function obtained employing mean-variance analysis also depends upon the covariance between relevance estimates of different documents. Once again, as no probability distribution over a document is considered, the covariance has scarce meaning. Covariance is then usually substituted by the correlation between documents term vectors.

In this paper, we revise the use of mean-variance analysis for ranking documents. In contrast with previous attempts, we instantiate the ranking strategy derived from the mean-variance analysis by considering a (discrete) relevance distribution associated to each retrieved document. This gives us the chance to actually compute the mean of the relevance estimates, their variances and the covariances between relevance estimates of documents. We show that this approach improves the performances of the retrieval system with respect to the ranking that would be obtained by considering a naïve ensemble of the mean of relevance estimates. This finding opens up new scenarios for the use of mean-variance analysis within Information Retrieval.

2 Mean-Variance Analysis Applied to Document Ranking

In [1,2], Wang et al. have shown how a document ranking strategy can be derived from the framework of mean-variance analysis. In particular, they suggest that a document should be placed at a particular rank position if this choice maximises the overall relevance (which is represented by the mean) of the ranked list at a given risk level (represented by the variance). We revise this paradigm, by assuming that a relevance distribution is assigned to each retrieved document. Given the relevance distribution of a document, indicated with $r_d(x) = r_d(1), \ldots, r_d(n)$, both the mean, $E[r_d]$, and the variance, $\mathrm{var}[r_d]$, can be computed as

$$E[r_d] = \sum_{i=1}^{n} \frac{r_d(i)}{n} \qquad\qquad \mathrm{var}[r_d] = E[r_d^2] - (E[r_d])^2$$

Furthermore, given two relevance distributions $r_d(x)$ and $r'_d(x)$, associated with documents d and d' respectively, the covariance between the two distribution is defined as

$$\mathrm{cov}[r_d, r_{d_i}] = E[r_d, r_{d_i}] - E[r_d]E[r_{d_i}]$$

Given these definitions, we can derive a ranking strategy over documents, similarly to the one proposed in [1,2]. In particular, if b is a parameter depending on the predilection (or aversion) of the user towards risk, and w_i a weight associated to the importance to the rank position i, documents can be ranked according to the mean-variance analysis paradigm. At each rank position, document d^* is selected if it satisfies the following condition:

$$d^* = \arg\max_d \left(E[r_d] - bw_n \,\mathrm{var}[r_d] - 2b \sum_{i=1}^{n-1} w_i \,\mathrm{cov}[r_d, r_{d_i}] \right) \tag{1}$$

Although the obtained formula appears to be equivalent to the one introduced in [1,2], it is fundamentally different. In fact, they (1) assume a point-wise relevance estimation, (2) conceptually substitute the variance in the estimations obtained for a document with the variance of the scores of the documents already ranked, and (3) approximate covariance between the relevance distributions of documents in terms of correlation between documents features. In the this paper, instead, the computations of mean and variance are referred to the (discrete) relevance distribution of a document, and similarly covariance is computed between distributions associated to different documents.

3 Experiments

We empirically tested the ranking approach based on mean-variance analysis outlined in the previous section. To do so, we employed the TREC Clueweb (part B) corpus and the topics used in the TREC 2009 Web track [5].
Deriving relevance distributions. In order to empirically investigate our revision of the mean-variance analysis paradigm for document ranking, a relevance distribution has to be derived for each retrieved document. We use the retrieval runs submitted to the TREC 2009 Web (ad-hoc) task (part B only). For each of the 50 queries used in TREC 2009 Web track, there were 34 rankings provided by as many retrieval systems (or variations of a common underlying system). Thus, for each document that has been retrieved by one of the participating systems, there are at maximum 34 relevance estimates. We used these estimates to form the relevance distribution for each document, after having normalised the sum of the scores of each document ranking to one. Note that a document might have been retrieved by a system, but not by any other. We thus discarded all the documents retrieved by only one system, as their variance would be null.
Experimental methodology. We compare our ranking approach (fed with the relevance distributions obtained from the TREC 2009 Web submissions) against the results obtained by a naïve strategy that re-ranks documents in decreasing order of the mean of their relevance estimations. We evaluate the obtained rankings in terms of MAP, MRR, precision at 10 (p@10), and α-NDCG, to assess both the ability of the proposed approach to improve the relevance and the diversity of the rankings.
Results and analysis. In table 1 we report the performances in terms of MAP, MRR and P@10 obtained by the compared approaches. Similarly, figure 1 shows performances in terms of α-NDCG at three different ranking depths. For the mean-variance analysis results, we let the parameter b vary in the range $[-300, 300]$, and we selected the run that provided best retrieval performances in terms of MAP and α-NDCG@10. The results evidence that when both traditional measures (with b set to -250) and diversity measures (with $b = -85$) are considered, the mean-variance analysis strategy provides better ranking than just considering the mean of the relevance estimates. Furthermore, the improvements are statistically significant (except when considering α-NDCG@5).

4 Conclusions

In this paper we have revisited the idea of using mean-variance analysis for document ranking, by considering the relevance distributions associated to documents retrieved by Information Retrieval systems. The outcome is a ranking function

Measure	*Baseline*	*Mean-Var*
MAP	0.0472	0.0533*
MRR	0.2194	0.3377*
p@10	0.1380	0.2000*

Table 1. Values of MAP, MRR and P@10 for the compared ranking approaches, when optimised for MAP. Statistical significance using t-test ($p < 0.05$) is indicated with *.

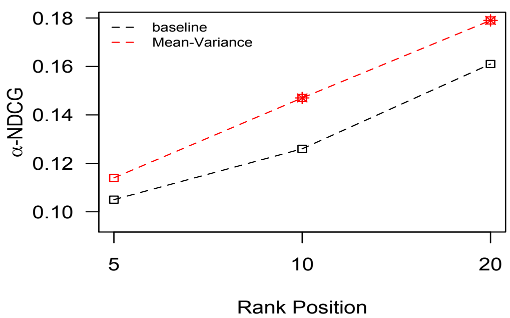

Figure 1. Values of α-NDCG at various rank positions for the compared ranking approaches, when optimised for α-NDCG@10. Statistical significance using t-test ($p < 0.05$) is indicated with *.

that provides a highly effective aggregation of document rankings. As a result, this approach could be applied to data fusion [6]. Further work will be directed towards: (1) investigating the robustness of the mean-variance approach under various conditions, and (2) comparing the proposed approach against state-of-the-art data fusion strategies.

Acknowledgments: The authors would like to thank the anonymous reviewers for their comments and P. Bruza, M. Sanderson, F. Scholer, and P. Thomas for useful discussions during early stages of this research. This work has been partially supported by the EPSRC Renaissance project (EP/F014384/1) and the Jim Gatheral Scholarship.

References

1. Wang, J.: Mean-Variance Analysis: A New Document Ranking Theory in Information Retrieval. In: ECIR ’09. (2009) 4–16
2. Wang, J., Zhu, J.: Portfolio Theory of Information Retrieval. In: SIGIR ’09. (2009) 115–122
3. Aly, R.B.N., Aiden, D., Hiemstra, D., Smeaton, A.: Beyond Shot Retrieval: Searching for Broadcast News Items using Language Models of Concepts. In: ECIR ’10. (2010) 241–252
4. Collins-Thompson, K.B.: Robust Model Estimation Methods for Information Retrieval. PhD thesis, Carnegie Mellon University, Pittsburgh, PA, USA (2008)
5. Clarke, C.L.A., Craswell, N., Soboroff, I.: Overview of the TREC 2009 web track. In: Proc. of TREC-2009. (2009)
6. Beitzel, S.M., Jensen, E.C., Chowdhury, A., Grossman, D., Frieder, O., Goharian, N.: Fusion of Effective Retrieval Strategies in the same Information Retrieval System. J. Am. Soc. Inf. Sci. Tech. **55**(10) (2004) 859–868