Theoretical Analysis of Interdependent Constraints in Pseudo-Relevance Feedback

Ali Montazeralghaem
Center for Intelligent Information Retrieval,
University of Massachusetts Amherst
montazer@cs.umass.edu

Hamed Zamani
Center for Intelligent Information Retrieval,
University of Massachusetts Amherst
zamani@cs.umass.edu

Azadeh Shakery
School of ECE, College of Engineering
University of Tehran, and
School of Computer Science, Institute for Research in Fundamental Sciences
shakery@ut.ac.ir

ABSTRACT

Axiomatic analysis is a well-defined theoretical framework for analytical evaluation of information retrieval models. The current studies in axiomatic analysis implicitly assume that the constraints (axioms) are independent. In this paper, we revisit this assumption and hypothesize that there might be interdependence relationships between the existing constraints. As a preliminary study, we focus on the pseudo-relevance feedback (PRF) models that have been theoretically studied using the axiomatic analysis approach. In this paper, we introduce two novel interdependent PRF constraints which emphasize on the effect of existing constraints on each other. We further modify two state-of-the-art PRF models, log-logistic and relevance models, in order to satisfy the proposed constraints. Experiments on three TREC newswire and web collections demonstrate that the proposed modifications significantly outperform the baselines, in all cases.

ACM Reference Format:
Ali Montazeralghaem, Hamed Zamani, and Azadeh Shakery. 2018. Theoretical Analysis of Interdependent Constraints in Pseudo-Relevance Feedback. In *SIGIR '18: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, July 8–12, 2018, Ann Arbor, MI, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3209978.3210156

1 INTRODUCTION

Many information retrieval (IR) models consist of heuristic components and/or make several simplifying assumptions that are not necessarily correct. Axiomatic analysis [1, 6, 11] provides a well-defined theoretic structure in order to study IR models analytically, which often lead to empirical improvements. In the axiomatic analysis framework, a number of constraints, also called axioms, are defined and IR models are designed or modified to satisfy these constraints. Previous work on axiomatic analysis literature assumes that axioms are independent of each other, and thus the models are studied given each axiom, separately.

In this paper, we revisit this assumption and hypothesize that there might be an interdependence relationship between different axioms. As a preliminary study, we focus on analyzing the pseudo-relevance feedback (PRF) models. PRF is a well-known strategy to address the vocabulary mismatch problem in information retrieval (IR). PRF assumes that the top retrieved documents in response to an initial query are relevant to the query and uses these documents for estimating a more accurate query model. Previous works have shown that PRF models can benefit from axiomatic analysis. For instance, Clinchant and Gaussier [4] proposed five constraints for PRF models and showed that even state-of-the-art PRF models do not satisfy all of the constraints. Following their work, a number of other constraints have been proposed and several modifications have been suggested for PRF models, e.g., see [2, 8–10].

In this paper, we propose two interdependent constraints. The first constraint takes the interdependence relationship of the term frequency (TF) and the inverse document frequency (IDF) into account. "TF effect" and "IDF effect", that are two existing constraints respectively corresponding to TF and IDF, have been previously considered independently [4]. Our first constraint shows the effect of IDF on the TF constraint. With a similar idea, our second constraint focuses on the relationship between term frequency and document relevance scores. In fact, this constraint indicates that the influence of term frequency on the feedback weight should depend on the relevance score of the documents containing the term.

Furthermore, we study two state-of-the-art PRF models, the log-logistic feedback model [3] and the relevance model [7], and modify these two models in order to satisfy the proposed interdependent constraints. The empirical evaluation on three TREC collections demonstrates the effectiveness of each of the proposed constraints. Our modifications to the log-logistic and relevance models lead to significant improvements in all collections. We believe that, our findings open up a new research direction by considering the interdependence relationship between constraints when studying IR models in different tasks.

2 METHODOLOGY

Several theoretic constraints have been proposed for pseudo-relevance feedback. Previous work [4, 8, 10] demonstrated that satisfying these constraints leads to significant improvements in retrieval performance. In this section, we propose two novel constraints for PRF models, which focus on the interdependence of the existing constraints. In the following subsections, we first propose two interdependent constraints and further modify two state-of-the-art PRF models in order to satisfy the proposed constraints.

A part of this work was done while Ali Montazeralghaem was a Master student at the University of Tehran.

SIGIR '18, July 8–12, 2018, Ann Arbor, MI, USA

ACM ISBN 978-1-4503-5657-2/18/07…$15.00
https://doi.org/10.1145/3209978.3210156

Notation. Let $FW(w; F, P_w, q)$ be a real-valued *feedback weight* function that assigns a weight to each candidate term w for a given query q. F and P_w denote the set of feedback documents for the query q and a set of term-dependent parameters, respectively. Let $RS(d, q)$ denote the relevance score of document d to query q and $SRS(w, F)$ denote the sum of relevance scores of the feedback documents containing the term w. For simplicity, we use $FW(w)$, $RS(d)$ and $SRS(w)$, hereafter. In the following equations, $tf(w, d)$ and $idf(w)$, respectively denote term frequency and inverse document frequency. The notation $|\cdot|$ represents the length of the given query/document or the size of the given set.

2.1 Interdependent Constraints

In this subsection, we introduce two interdependent constraints for pseudo-relevance feedback models.

[TF-IDF effect] Let w_1 and w_2 be two vocabulary terms, such that $idf(w_1) > idf(w_2)$ and $SRS(w_1) = SRS(w_2)$. Assume that there exists a document $d \in F$, where $tf(w_1, d) = tf(w_2, d) > 0$ and $FW(w_1; F \backslash \{d\}) = FW(w_2; F \backslash \{d\})$. In this case, if we add both terms to this document such that $tf'(w_1, d) = tf'(w_2, d) = tf(w_1, d) + 1$ and $FW'(w_1)$ and $FW'(w_1)$ be updated weights, then we will have:

$$FW'(w_1) - FW'(w_2) > FW(w_1) - FW(w_2) \tag{1}$$

Formally writing, for all candidate feedback terms w, the following constraints should be satisfied:

$$\frac{\partial^2 FW(w)}{\partial tf(w, d).\partial idf(w)} > 0$$

The intuition behind this constraint is that increasing the term frequency of common terms (low idf) should have less impact on the feedback weight, compared to the rare terms (high idf). From the information theory perspective, if we add a rare term to a document, we provide more information, compared to the scenario of adding a common term. Consequently, the difference of feedback weight between two terms caused by discrimination value of them, should increase when these two terms are added to the feedback document. This shows the interdependence between term frequency and inverse document frequency that have been considered as independent axioms in prior work [4].

[TF-SRS effect] Let w_1 and w_2 be two vocabulary terms, such that $idf(w_1) = idf(w_2)$ and $SRS(w_1) > SRS(w_2)$. Let $d \in F$ be a feedback document, where $tf(w_1, d) = tf(w_2, d) > 0$. In this case, if we add both terms to the document, i.e., $tf'(w_1, d) = tf'(w_2, d) = tf(w_1, d) + 1$, then we will have:

$$FW'(w_1) - FW'(w_2) > FW(w_1) - FW(w_2) \tag{2}$$

which can be formally formulated as:

$$\frac{\partial^2 FW(w)}{\partial tf(w, d).\partial SRS(w)} > 0$$

where w is a candidate feedback term. In other words, according to the relevance effect constraint [2, 10], the terms appearing in the documents with higher relevance scores should get higher feedback weights. The TF-SRS constraint indicates that increase in frequency of the words with higher $SRS(w)$ values should lead to higher feedback weights. The intuition behind this idea can be mapped to the one proposed for the previous constraint.

2.2 Modifying the Log-Logistic Model

The log-logistic (LL) feedback model [3] is a state-of-the-art PRF model that has been shown to outperform several PRF models, including the geometric relevance model [12] and the mixture model [13]. The feedback weight function in the log-logistic model is defined as follows:

$$FW_{LL}(w) = \frac{1}{|F|} \sum_{d \in F} FW_{LL}(w, d) = \frac{1}{|F|} \sum_{d \in F} \log(\frac{t(w, d) + \lambda_w}{\lambda_w}) \tag{3}$$

where $\lambda_w = \frac{N_w}{N}$ is the fraction of the documents that contain the term w in the whole collection. Also, $t(w, d) = tf(w, d) \log(1 + c\frac{avg_l}{|d|})$ represents a term frequency function normalized by the document length, where avg_l denotes the average document length and c is a free hyper-parameter. A modification of the log-logistic model has been proposed recently [8], called LLR, in order to satisfy the "relevance effect" constraint [10]:

$$FW_{LLR}(w) = \frac{1}{|F|} \sum_{d \in F} FW_{LL}(w, d) * RS(Q, d) \tag{4}$$

where RS denotes the relevance score function. We focus on this model since it significantly outperforms the original log-logistic model [8]. This model satisfies "TF-IDF effect" because:

$$\frac{\partial^2 FW_{LLR}(w)}{\partial tf(w, d).\partial idf(w)} = \frac{\log(1 + c\frac{avg_l}{|d|})}{(\frac{1}{\lambda_w} \log(1 + c\frac{avg_l}{|d|}) tf(w, d) + 1)^2} > 0 \tag{5}$$

Although this approach satisfies the first proposed constraint, there is only a light interdependence relationship between tf and idf. In other words, the derivative of the LLR weight function with respect to $tf(w, d)$, which is given as follows, does not have a strong dependence to idf:

$$\frac{\partial FW_{LLR}(w)}{\partial tf(w, d)} = \frac{\frac{1}{\lambda_w} \log(1 + c\frac{avg_l}{|d|})}{\frac{1}{\lambda_w} \log(1 + c\frac{avg_l}{|d|}) tf(w, d) + 1} \tag{6}$$

As shown in the above equation, if we omit 1 from the denominator, the derivative becomes independent from idf. We believe that increasing this derivative with respect to the idf could lead to higher performance. To overcome this issue, we re-write this function as follows:

$$FW_{TF-IDF}(w) = \frac{1}{|F|} \sum_{d \in F} \log(\frac{t(w, d)^{A(w)} + \lambda_w}{\lambda_w}) \tag{7}$$

where

$$A(w) = \log(\frac{1}{\lambda_w}) = \log(\frac{N}{N_w}) \tag{8}$$

Now the derivative of this function is:

$$\frac{\partial FW_{TF-IDF}(w)}{\partial tf(w, d)} = \frac{\frac{1}{\lambda_w} \log(1 + c\frac{avg_l}{|d|}) A(w) tf(w, d)^{(A(w)-1)}}{\frac{1}{\lambda_w} \log(1 + c\frac{avg_l}{|d|}) tf(w, d)^{A(w)} + 1} \tag{9}$$

According to the above equation, the derivative of FW_{TF-IDF} is more dependent to idf, and when we omit 1 from the denominator the derivative nearly will be $A(w)$ that is stronger from Equation (6) for satisfying TF-IDF constraint. Figure 1 illustrates the correlation between idf and derivative of LLR and LLR+TF-IDF with respect to term frequency (Equations (6) and (9)). Based on this plot, LLR weight function has a light interdependence relationship between tf and idf whereas LLR+TF-IDF improves it.

It can be shown that $\frac{\partial^2 FW_{LLR}(w)}{\partial tf(w,d).\partial SRS(w)} = 0$, which means that the log-logistic models (both LL and LLR) do not satisfy the "TF-SRS effect" constraint. Hence, we modify the weight function as follows in order to satisfy this constraint:

$$FW_{TF-SRS}(w) = Com(w,F)\frac{1}{|F|}\sum_{d\in F}\log(\frac{t(w,d)+\lambda_w}{\lambda_w}) \quad (10)$$

where $Com(w,F)$ is defined as follows:

$$Com(w,F) = \frac{\sum_{d\in F} RS(d) * I(w,d)}{\sum_{d\in F} RS(d)} \quad (11)$$

where $I(w,d)$ denotes the occurrence of term w in document d. Hence, the modified log-logistic model is:

$$FW_{All}(w) = Com(w,F)\frac{1}{|F|}\sum_{d\in F}\log(\frac{t(w,d)^{A(w)}+\lambda_w}{\lambda_w}) \quad (12)$$

2.3 Modifying the Relevance Model

In this subsection, we focus on RM3, a well-known and state-of-the-art variant of the relevance models proposed by Lavrenko and Croft [7]. The feedback weight of each term w for a given query Q in RM3 is computed as follows:

$$FW_{RM3}(w) = p(w|Q) = \frac{p(w,Q)}{p(Q)} \propto p(w,Q) = p(w,q_1,...,q_k) \quad (13)$$

where q_i denotes the i^{th} query term.

Clinchant and Gaussier [4] showed that relevance model does not satisfy IDF effect, and this means that this model cannot satisfy TF-IDF effect. This model also does not have any component of SRS and therefore cannot satisfy TF-SRS effect.

We re-write this function by bringing the feedback set into the feedback weight formula as follows:

$$\begin{aligned} p(w|Q,F) \propto p(w,Q|F) &= p(w,q_1,...,q_k|F) \\ &= \sum_{D\in F} p(D)p(w,q_1,...,q_k|D,F) \end{aligned} \quad (14)$$

The conditional independence assumption of terms leads to the following equation:

$$p(w,q_1,...,q_k|D,F) = p(w|D,F)\prod_{i=1}^{k} p(q_i|D,F) \quad (15)$$

We estimate each of the conditional as follows:

$$p(w|D,F) = \frac{p(D|w,F)p(w|F)p(F)}{p(D|F)p(F)} \quad (16)$$

Under the assumption of independence of a feedback document given a term from the whole feedback set, we have:

$$p(w|D,F) = \frac{p(D|w)p(w|F)}{\sum_{w_i\in D} p(D|w_i)p(w_i|F)} \quad (17)$$

where $p(D|w)$ can be computed by Bayes rule $p(D|w) = \frac{p(w|D)p(D)}{p(w)}$ where we consider $p(D)$ uniform for all documents and $p(w|D)$ is computed using the maximum likelihood estimation: $p(w|D) = \frac{tf(w,D)}{|D|}$.

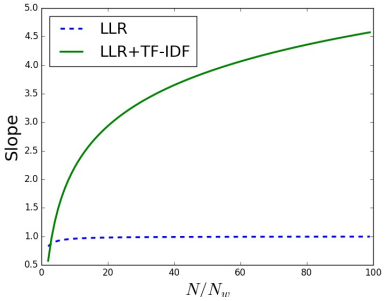

Figure 1: Correlation between idf and the derivative of LLR and LLR+TF-IDF formulas with respect to TF (Eq. (6) and (9)).

To satisfy the proposed constraints, we estimate $p(w|F)$ as follows:

$$p(w|F) \propto \sum_{D\in F} RS(D) * I(w,D)$$

which shows how this term appears in the relevant documents. $p(w)$ can also be estimated using the document frequency of the term in the whole collection, which indicates how common the term is. Other computation details are the same as the original relevance models [7].

It can be proved that the proposed modification to RM3 satisfies both TF-IDF and TF-SRS constraints as well as the IDF effect constraint proposed in [4]. We call this model RM3+ALL.

3 EXPERIMENTS

3.1 Collections and Experimental Setup

We used three standard TREC collections in our experiments: AP (Associated Press 1988-89, TREC topics 51-200), Robust (TREC Robust Track 2004 collection, TREC topics 301-450 & 601-700) and WT10g (TREC Web Track 2001-2002, TREC topics 451-550). The first two collections are homogeneous collections containing news articles, while the third collection is a heterogeneous collection containing web pages. The WT10g collection is noisier than the newswire collections.

All documents are stemmed using the Porter stemmer and stopped using the standard INQUERY stopword list. [1]

3.1.1 Parameter Setting. The number of feedback documents, the number of feedback terms, the feedback coefficient and the parameter c in the Log-logistic model are set using 2-fold cross-validation over the queries of each collection. We sweeped the number of feedback documents between $\{10, 25, 50, 75, 100\}$, feedback terms between $\{50, 100, \cdots, 300\}$, the feedback coefficient between $\{0, 0.1, \cdots, 1\}$, and the parameters c between $\{2, 4, 6, 8, 10\}$.

3.1.2 Evaluation Metrics. We use mean average precision (MAP) of the 1000 top-ranked documents as the main metric to evaluate the retrieval effectiveness. We also report the precision of the top 10 retrieved documents (P@10). Furthermore, we consider the robustness index (RI) [5] to evaluate the robustness of methods. Statistically significant differences of performances are determined using the two-tailed paired t-test computed at a 95% confidence level over average precision per query.

3.2 Results and Discussion

3.2.1 Evaluating the Modified Log-Logistic Model. Baselines: (1) the document retrieval model without feedback (NoPRF) computed

[1] The experiments were carried out using the Lemur toolkit (http://lemurproject.org/)

Table 1: Performance of the proposed modifications and the baselines. Superscripts 0/1/2/3/4/6 denote that the MAP improvements over NoPRF/LL/LLR/LLR+TF-IDF/LLR+TF-SRS/RM3 are statistically significant.

Method	AP			Robust			WT10g		
	MAP	P@10	RI	MAP	P@10	RI	MAP	P@10	RI
NoPRF	0.2642	0.4260	–	0.2490	0.4237	–	0.2080	0.3030	–
LL	0.3379^{0}	0.4648	0.17	0.2816^{0}	0.4385	0.32	0.2127	0.3156	0.11
LLR	0.3409^{0}	0.4668	0.19	$0.2909^{0,1}$	0.4442	0.36	$0.2299^{0,1}$	0.3339	0.19
LLR+TF-IDF	$0.3453^{0,1,2}$	0.4675	0.20	$0.2956^{0,1,2}$	0.4490	0.37	$0.2370^{0,1}$	0.3269	0.19
LLR+TF-SRS	$0.3470^{0,1,2}$	0.4702	0.21	$0.2965^{0,1,2}$	0.4418	0.39	$0.2371^{0,1}$	0.3298	0.24
LLR+ALL	$0.3490^{0,1,2,4}$	0.4735	0.20	$0.2986^{0,1,2}$	0.4478	0.37	$0.2401^{0,1,2}$	0.3359	0.17
RM3	0.3392^{0}	0.4561	0.17	0.2919^{0}	0.4322	0.24	0.2213^{0}	0.3166	0.12
RM3+ALL	$0.3460^{0,6}$	0.4695	0.19	$0.2967^{0,6}$	0.4462	0.25	$0.2297^{0,6}$	0.3177	0.20

by query likelihood, (2) the original log-logistic feedback model (LL) [3], and (3) the modified log-logistic model (LLR) [8] that satisfies the relevance effect constraint.

To study the effect of each constraint in the retrieval performance, we modify the log-logistic model based on each constraint, separately. Equations (7) and (10) are the feedback weight functions that satisfy the TF-IDF and the TF-SRS constraints, respectively. We also modify the log-logistic model by considering all of these constraints, called LL+ALL (see Equation (12)). The results obtained by the baselines and those achieved by the proposed modifications are reported in Table 1. LL outperforms the NoPRF baseline in all cases, which indicates the effectiveness of the log-logistic model and validates the findings presented in prior work [3, 8]. Both LLR+TF-IDF and LLR+TF-SRS outperform all three baselines (NoPRF, LL, and LLR), which demonstrates the effectiveness of the proposed constraints. The improvements achieved by LLR+TF-SRS is higher than those obtained by LLR+TF-IDF. This indicates that the interdependence between tf and SRS which models the local importance of the term is more effective than the interaction between tf and idf which models the global importance of the term. In addition, LLR+TF-SRS and LLR+TF-IDF are shown to be more robust than the baselines in all the collections. LLR+TF-SRS performs more robust compared to LLR+TF-IDF, in all the collections, especially the web collection. The MAP improvements for both LLR+TF-IDF and LLR+TF-SRS methods are close to each other in all the collections specially in the noisy collection (WT10g) which shows that both of these two constraints are important for PRF methods.

LLR+All that satisfies both constraints outperforms the baselines. These improvements are statistically significant in all cases.

3.2.2 Evaluating the Modified Relevance Model. In this set of experiments, we consider two baselines: (1) the document retrieval method without feedback (NoPRF) computed by query likelihood, and (2) the RM3 method [7]. Since the proposed approach satisfies both constraints simultaneously (we do not propose different solutions for satisfying different axioms), we only report the results for RM3+ALL. As shown in Table 1, RM3+ALL significantly outperforms both baselines (NoPRF and RM3) in all collections, in terms of MAP. The P@10 values achieved by the proposed method is also higher than those achieved by the baselines, in all cases. This shows that our modification to RM3 leads to better performance by satisfying the proposed constraints. Our modification also makes the relevance model more robust, especially in the web collection.

4 CONCLUSIONS AND FUTURE WORK

In this paper, we proposed two novel constraints for pseudo-relevance feedback models, which focus on the interdependence relation of the existing constraints that have been previously considered as independent. To show the importance of these interdependent constraints, we studied and modified two state-of-the-art pseudo-relevance feedback models; the log-logistic and the relevance models. Our evaluation on three standard newswire and web collections investigates the effect of each of these constraints on the overall effectiveness and robustness of the models. Our modifications lead to significant improvements over the baselines in all the collections. These observations suggest that taking the interdependence relationship of the constraints into account might lead to designing more accurate IR models, which could be an interesting research direction for future work.

Acknowledgements. This work was supported in part by the Center for Intelligent Information Retrieval. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the sponsor.

REFERENCES

[1] M. Ariannezhad, A. Montazeralghaem, H. Zamani, and A. Shakery. Improving retrieval performance for verbose queries via axiomatic analysis of term discrimination heuristic. In *SIGIR '17*, pages 1201–1204. ACM, 2017.
[2] M. Ariannezhad, A. Montazeralghaem, H. Zamani, and A. Shakery. Iterative Estimation of Document Relevance Score for Pseudo-Relevance Feedback. In *ECIR '17*, 2017.
[3] S. Clinchant and E. Gaussier. Information-based Models for Ad Hoc IR. In *SIGIR '10*, pages 234–241, 2010.
[4] S. Clinchant and E. Gaussier. A Theoretical Analysis of Pseudo-Relevance Feedback Models. In *ICTIR '13*, pages 6–13, 2013.
[5] K. Collins-Thompson. Reducing the Risk of Query Expansion via Robust Constrained Optimization. In *CIKM '09*, pages 837–846, 2009.
[6] H. Fang, T. Tao, and C. Zhai. A formal study of information retrieval heuristics. In *SIGIR '04*, pages 49–56, 2004.
[7] V. Lavrenko and W. B. Croft. Relevance Based Language Models. In *SIGIR '01*, pages 120–127, 2001.
[8] A. Montazeralghaem, H. Zamani, and A. Shakery. Axiomatic Analysis for Improving the Log-Logistic Feedback Model. In *SIGIR '16*, pages 765–768, 2016.
[9] A. Montazeralghaem, H. Zamani, and A. Shakery. Term proximity constraints for pseudo-relevance feedback. In *SIGIR '17*, pages 1085–1088. ACM, 2017.
[10] D. Pal, M. Mitra, and S. Bhattacharya. Improving Pseudo Relevance Feedback in the Divergence from Randomness Model. In *ICTIR '15*, pages 325–328, 2015.
[11] R. Rahimi, A. Shakery, and I. King. Axiomatic analysis of cross-language information retrieval. In *CIKM '14*, pages 1875–1878. ACM, 2014.
[12] J. Seo and W. B. Croft. Geometric Representations for Multiple Documents. In *SIGIR '10*, pages 251–258, 2010.
[13] C. Zhai and J. Lafferty. Model-based Feedback in the Language Modeling Approach to Information Retrieval. In *CIKM '01*, pages 403–410, 2001.