

More Languages, More MAP?: A Study of Multiple Assisting Languages in Multilingual PRF

Vishal Vachhani Manoj K. Chinnakotla Mitesh M. Khapra Pushpak Bhattacharyya

Department of Computer Science and Engineering,
Indian Institute of Technology Bombay

{vishalv,manoj,miteshk,pb}@cse.iitb.ac.in

Abstract

Multilingual Pseudo-Relevance Feedback (MultiPRF) is a framework to improve the PRF of a *source language* by taking the help of another language called *assisting language*. In this paper, we extend the MultiPRF framework to include multiple assisting languages. We consider three different configurations to incorporate multiple assisting languages - a) Parallel - all assisting languages combined simultaneously b) Serial - assisting languages combined in sequence one after another and c) Selective - dynamically selecting the best feedback model for each query. We study their effect on MultiPRF performance. Results using multiple assisting languages are mixed and it helps in boosting MultiPRF accuracy only in some cases. We also observe that MultiPRF becomes more robust with increase in number of assisting languages.

1 Introduction

Pseudo-Relevance Feedback (PRF) (Buckley et al., 1994; Xu and Croft, 2000; Mitra et al., 1998) is known to be an effective technique to improve the effectiveness of Information Retrieval (IR) systems. In PRF, the top '*k*' documents from the ranked list retrieved using the initial keyword query are *assumed to be relevant*. Later, these documents are used to refine the user query and the final ranked list is obtained using the above refined query. Although PRF has been shown to improve retrieval, it suffers from the following drawbacks: (a) *Lexical and Semantic Non-Inclusion*: the type of term associations ob-

tained for query expansion is restricted to only co-occurrence based relationships in the feedback documents and (b) *Lack of Robustness*: due to the inherent assumption in PRF, *i.e.*, relevance of top *k* documents, performance is sensitive to that of the initial retrieval algorithm and as a result is not robust. Typically, larger coverage ensures higher proportion of relevant documents in the top *k* retrieval (Hawking et al., 1999). However, some resource-constrained languages do not have adequate information coverage in their own language. For example, languages like Hungarian and Finnish have meager online content in their own languages.

Multilingual Pseudo-Relevance Feedback (MultiPRF) (Chinnakotla et al., 2010a) is a novel framework for PRF to overcome the above limitations of PRF. It does so by taking the help of a different language called the *assisting language*. Thus, the performance of a resource-constrained language could be improved by harnessing the good coverage of another language. MultiPRF showed significant improvements on standard CLEF collections (Braschler and Peters, 2004) over state-of-art PRF system. On the web, each language has its own exclusive topical coverage besides sharing a large number of common topics with other languages. For example, information about Saudi Arabia government policies and regulations is more likely to be found in Arabic language web and also information about a local event in Spain is more likely to be covered in Spanish web than in English. Hence, using multiple languages in conjunction is more likely to ensure satisfaction of the user information need and hence will be more robust.

In this paper, we extend the MultiPRF framework to multiple assisting languages. We study

the various possible ways of combining the models learned from multiple assisting languages. We propose three different configurations for including multiple assisting languages in MultiPRF - a) Parallel b) Serial and c) Selective. In Parallel combination, all the assisting languages are combined simultaneously using interpolation. In Serial configuration, the assisting languages are applied in sequence one after another and finally, in Selective configuration, the best feedback model is dynamically chosen for each query. We experiment with each of the above configurations and present both quantitative and qualitative analysis of the results. Results using multiple assisting languages are mixed and it helps in boosting MultiPRF accuracy only in some cases. We also observe that MultiPRF becomes more robust with increase in number of assisting languages. Besides, we also study the relation between number of assisting languages, coverage and the MultiPRF accuracy.

The paper is organized as follows: Section 2, explains the Language Modeling (LM) based PRF approach. Section 3, describes the MultiPRF approach. Section 4 explains the various configurations to extend MultiPRF for multiple assisting languages. Section 6 presents the results and discussions. Finally, Section 7 concludes the paper.

2 PRF in the LM Framework

The Language Modeling (LM) Framework allows PRF to be modeled in a principled manner. In the LM approach, documents and queries are modeled using multinomial distribution over words called *document language model* $P(w|D)$ and *query language model* $P(w|\Theta_Q)$ respectively. For a given query, the document language models are ranked based on their proximity to the query language model, measured using KL-Divergence.

$$KL(\Theta_Q||D) = \sum_w P(w|\Theta_Q) \cdot \log \frac{P(w|\Theta_Q)}{P(w|D)}$$

Since the query length is short, it is difficult to estimate Θ_Q accurately using the query alone. In PRF, the top k documents obtained through the initial ranking algorithm are assumed to be relevant and used as feedback for improving the estimation of Θ_Q . The feedback documents contain both relevant and noisy terms from which

Symbol	Description
Θ_Q	Query Language Model
$\Theta_{L_1}^F$	Feedback Language Model obtained from PRF in L_1
$\Theta_{L_2}^F$	Feedback Language Model obtained from PRF in L_2
$\Theta_{L_1}^{Trans}$	Feedback Model Translated from L_2 to L_1
$t(f e)$	Probabilistic Bi-Lingual Dictionary from L_2 to L_1
β, γ	Interpolation coefficients used in MultiPRF

Table 1: Glossary of Symbols used in explaining MultiPRF the feedback language model is inferred based on a Generative Mixture Model (Zhai and Lafferty, 2001).

Let $D_F = \{d_1, d_2, \dots, d_k\}$ be the top k documents retrieved using the initial ranking algorithm. Zhai and Lafferty (Zhai and Lafferty, 2001) model the feedback document set D_F as a mixture of two distributions: (a) the *feedback language model* and (b) the *collection model* $P(w|C)$. The feedback language model is inferred using the EM Algorithm (Dempster et al., 1977), which iteratively accumulates probability mass on the most *distinguishing* terms, *i.e.* terms which are more frequent in the feedback document set than in the entire collection. To maintain query focus the final converged feedback model, Θ_F is interpolated with the initial query model Θ_Q to obtain the final query model Θ_{Final} .

$$\Theta_{Final} = (1 - \alpha) \cdot \Theta_Q + \alpha \cdot \Theta_F$$

Θ_{Final} is used to re-rank the corpus using the KL-Divergence ranking function to obtain the final ranked list of documents. Henceforth, we refer to the above technique as *Model Based Feedback (MBF)*.

3 Multilingual Pseudo-Relevance Feedback (MultiPRF)

Chinnakotla et al. (Chinnakotla et al., 2010a; Chinnakotla et al., 2010b) propose the MultiPRF approach which overcomes the fundamental limitations of PRF with the help of an assisting collection in a different language. Given a query Q in the source language L_1 , it is automatically translated into the assisting language L_2 . The documents in the L_2 collection are ranked using the query likelihood ranking function (John Lafferty and Chengxiang Zhai, 2003). Using the top k documents, they estimate the feedback model using MBF as described in the previous section. Similarly, they also estimate a feedback model using

the original query and the top k documents retrieved from the initial ranking in L_1 . Let the resultant feedback models be $\Theta_{L_2}^F$ and $\Theta_{L_1}^F$ respectively. The feedback model estimated in the assisting language $\Theta_{L_2}^F$ is translated back into language L_1 using a probabilistic bi-lingual dictionary $t(f|e)$ from $L_2 \rightarrow L_1$ as follows:

$$P(f|\Theta_{L_1}^{Trans}) = \sum_{\forall e \in L_2} t(f|e) \cdot P(e|\Theta_{L_2}^F) \quad (1)$$

The probabilistic bi-lingual dictionary $t(f|e)$ is learned from a parallel sentence-aligned corpora in $L_1 - L_2$ based on word level alignments. The probabilistic bi-lingual dictionary acts as a rich source of morphologically and semantically related feedback terms. Thus, the translation model adds related terms in L_1 which have their source as the term from feedback model $\Theta_{L_2}^F$. The final MultiPRF model is obtained by interpolating the above translated feedback model with the original query model and the feedback model of language L_1 as given below:

$$\Theta_{L_1}^{Multi} = (1 - \beta - \gamma) \cdot \Theta_Q + \beta \cdot \Theta_{L_1}^F + \gamma \cdot \Theta_{L_1}^{Trans} \quad (2)$$

In order to retain the query focus during back translation, the feedback model in L_2 is interpolated with the translated query before translation of the L_2 feedback model. The parameters β and γ control the relative importance of the original query model, feedback model of L_1 and the translated feedback model obtained from L_1 and are tuned based on the choice of L_1 and L_2 .

4 Extending MultiPRF to Multiple Assisting Languages

In this section, we extend the MultiPRF model described earlier to multiple assisting languages. Since each language produces a different feedback model, there could be different ways of combining these models as suggested below.

Parallel: One way is to include the new assisting language model using one more interpolation coefficient which gives the effect of using multiple assisting languages in *parallel*.

Serial: Alternately, we can have a *serial* combination wherein language L_2 is first assisted

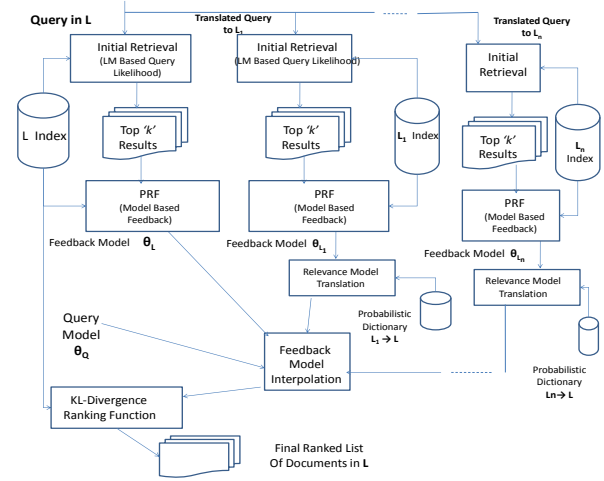


Figure 1: Schematic of the Multilingual PRF Approach Using Parallel Assistance

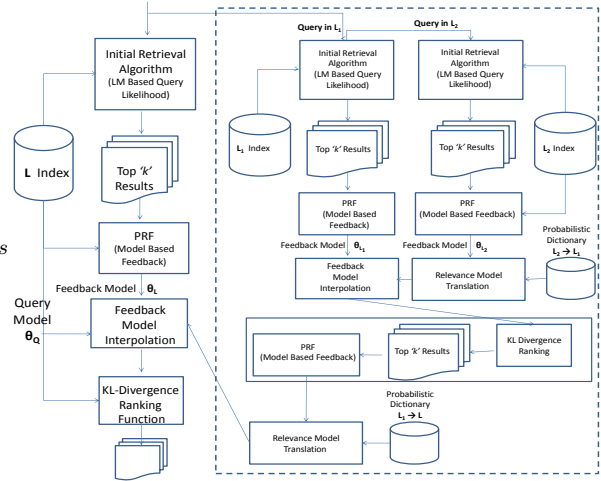


Figure 2: Schematic of the Multilingual PRF Approach Using Serial Assistance

by language L_3 and then this MultiPRF system is used to assist the source language L_1 .

Selective: Finally, we can have selective assistance wherein we dynamically select which assisting language to use based on the input query.

Below we describe each of these systems in detail.

4.1 Parallel Combination

The MultiPRF model as explained in section 3 interpolates the query model of L_1 with the MBF of L_1 and the translated feedback model of the assisting language L_2 . The most natural extension to this approach is to translate the query into multiple languages instead of a single language and collect the feedback terms from the initial re-

Language	CLEF Collection Identifier	Description	No. of Documents	No. of Unique Terms	CLEF Topics (No. of Topics)
English	EN-02+03	LA Times 94, Glasgow Herald 95	169477	234083	91-200 (67)
French	FR-02+03	Le Monde 94, French SDA 94-95	129806	182214	91-200 (67)
German	DE-02+03	Frankfurter Rundschau 94, Der Spiegel 94-95, German SDA 94-95	294809	867072	91-200 (67)
Finnish	FI-02+03	Aamulehti 94-95	55344	531160	91-200 (67)
Dutch	NL-02+03	NRC Handelsblad 94-95, Algemeen Dagblad 94-95	190604	575582	91-200 (67)
Spanish	ES-02+03	EFE 94, EFE 95	454045	340250	91-200 (67)

Table 2: Details of the CLEF Datasets used for Evaluating the MultiPRF approach. The number shown in brackets of the final column CLEF Topics indicate the actual number of topics used during evaluation.

trieval of each of these languages. The translated feedback models resulting from each of these retrievals can then be interpolated to get the final parallel MultiPRF model. Specifically, if L_1 is the source language and L_2, L_3, \dots, L_n are assisting languages then final parallel MultiPRF model can be obtained by generalizing Equation 2 as shown below:

$$\Theta_{L_1}^{MultiAssist} = (1 - \beta - \sum_i \alpha_i) \cdot \Theta_Q + \beta \cdot \Theta_F + \sum_i \alpha_i \cdot \Theta_{L_i}^{Trans} \quad (3)$$

The schematic representation of parallel combination is shown in Figure 1.

4.2 Serial Combination

Let L_1 be the source language and let L_2 and L_3 be two assisting languages. A serial combination can then be achieved by cascading two MultiPRF systems as described below:

1. Construct a MultiPRF system with L_2 as the source language and L_3 as the assisting language. We call this system as L_2L_3 -MultiPRF system.
2. Next, construct a MultiPRF system with L_1 as the source language and L_2L_3 -MultiPRF as the assisting system.

As compared to a single assistance system where only L_2 is used as the assisting language for L_1 , here the performance of language L_2 is first boosted using L_3 as the assisting language. This boosted system is then used for assisting L_1 . Also note that unlike parallel assistance here we do not introduce an extra interpolation co-efficient in the original MultiPRF model given in Equation 2. The schematic representation of serial combination is shown in Figure 2.

4.3 Selective Assistance

We motivate selective assistance by posing the following question: “Given a source language L_1 and two assisting languages L_2 and L_3 , is it possible that L_2 is ideal for assisting some queries whereas L_3 is ideal for assisting some other queries?” For example, suppose L_2 has a rich collection of TOURISM documents whereas L_3 has a rich collection of HEALTH documents. Now, given a query pertaining to TOURISM domain one might expect L_2 to serve as a better assisting language whereas given a query pertaining to the HEALTH domain one might expect L_3 to serve as a better assisting language. This intuition can be captured by suitably changing the interpolation model as shown below:

$$\Theta_L^{Best} = SelectBestModel(\Theta_L^F, \Theta_{L_1}^{Trans}, \Theta_{L_2}^{Trans}, \Theta_{L_{12}}^{Trans})$$

$$\Theta_{L_1}^{Multi} = (1 - \alpha) \cdot \Theta_Q + \alpha \cdot \Theta_L^{Best} \quad (4)$$

where, $SelectBestModel()$ gives the best model for a particular query using the algorithm mentioned below which is based on minimizing the query drift as described in (?):

1. Obtain the four feedback models, viz., $\Theta_L^F, \Theta_{L_1}^{Trans}, \Theta_{L_2}^{Trans}, \Theta_{L_{12}}^{Trans}$
2. Build a language model (say, LM) using query Q and top -100 documents of initial retrieval in language L .
3. Find the KL-Divergence between LM and the four models obtained during step 1.
4. Select the model which has minimum KL-Divergence score from LM . Call this model Θ_L^{Best} .
5. Get the final model by interpolating the query model, Θ_Q , with Θ_L^{Best} .

5 Experimental Setup

We evaluate the performance of our system using the standard CLEF evaluation data in six languages, widely varying in their familial relationships - Dutch, German, English, French, Spanish and Finnish. The details of the collections and their corresponding topics used for MultiPRF are given in Table 2. Note that, in each experiment, we choose assisting collections such that the topics in the source language are covered in the assisting collection so as to get meaningful feedback terms. In all the topics, we only use the *title* field. We ignore the topics which have no relevant documents as the true performance on those topics cannot be evaluated.

We use the Terrier IR platform (Ounis et al., 2005) for indexing the documents. We perform standard tokenization, stop word removal and stemming. We use the Porter Stemmer for English and the stemmers available through the Snowball package for other languages. Other than these, we do not perform any language-specific processing on the languages. In case of French, since some function words like *l', d' etc.*, occur as prefixes to a word, we strip them off during indexing and query processing, since it significantly improves the baseline performance. We use standard evaluation measures like *MAP*, *P@5* and *P@10* for evaluation. Additionally, for assessing robustness, we use the Geometric Mean Average Precision (GMAP) metric (Robertson, 2006) which is also used in the TREC Robust Track (Voorhees, 2006). The probabilistic bi-lingual dictionary used in MultiPRF was learnt automatically by running GIZA++: a word alignment tool (Och and Ney, 2003) on a parallel sentence aligned corpora. For all the above language pairs we used the *Europarl Corpus* (Philipp, 2005). We use Google Translate as the query translation system as it has been shown to perform well for the task (Wu et al., 2008). We use two-stage Dirichlet smoothing with the optimal parameters tuned based on the collection (Zhai and Lafferty, 2004). We tune the parameters of MBF, specifically λ and α , and choose the values which give the optimal performance on a given collection. We observe that the optimal parameters γ and β are uniform across collections and vary in the range 0.4-0.48. We

Source Langs	Assist. Langs		MBF	MultiPRF (L_1)	MultiPRF (L_2)	MultiPRF (L_1, L_2)
EN	DE-NL	MAP	0.4495	0.4464	0.4471	0.4885(4.8) [†]
		P@5	0.4955	0.4925	0.5045	0.5164(2.4)
		P@10	0.4328	0.4343	0.4373	0.4463(2.1)
	DE-FI	MAP	0.4495	0.4464	0.4545	0.4713(3.7) [†]
		P@5	0.4955	0.4925	0.5194	0.5224(1.2)
		P@10	0.4328	0.4343	0.4373	0.4507(3.1)
	NL-ES	MAP	0.4495	0.4471	0.4566	0.4757(4.2) [†]
		P@5	0.4955	0.5045	0.5164	0.5224(0.6)
		P@10	0.4328	0.4373	0.4537	0.4448(2.4)
	ES-FR	MAP	0.4495	0.4566	0.4563	0.48(5.1) [†]
		P@5	0.4955	0.5164	0.5075	0.5224(1.2)
		P@10	0.4328	0.4537	0.4343	0.4388(-3.3)
	ES-FI	MAP	0.4495	0.4566	0.4545	0.48(5.1) [†]
		P@5	0.4955	0.5164	0.5194	0.5254(1.7)
		P@10	0.4328	0.4537	0.4373	0.4403(-3.0)
	FR-FI	MAP	0.4495	0.4563	0.4545	0.4774(4.6)
		P@5	0.4955	0.5075	0.5194	0.5284(4.1) [†]
		P@10	0.4328	0.4343	0.4373	0.4373(0.7)
EN-FR	EN-FR	MAP	0.3578	0.3411	0.3553	0.3688(3.8)
		P@5	0.3821	0.394	0.397	0.4149(4.5) [†]
		P@10	0.3105	0.3463	0.3433	0.3433(0.1)
	NL-DE	MAP	0.3578	0.3722	0.3796	0.3929(3.5)
		P@5	0.3821	0.406	0.403	0.4149(3.0)
		P@10	0.3105	0.3478	0.3582	0.3597(0.4)
ES-DE	MAP	0.3578	0.369	0.3796	0.4058(6.9) [†]	
	P@5	0.3821	0.4119	0.403	0.4239(5.2)	
	P@10	0.3105	0.3448	0.3582	0.3612(0.8)	
FI	FR-DE	MAP	0.3578	0.3553	0.3796	0.3988(5.1) [†]
		P@5	0.3821	0.397	0.403	0.406(0.7)
		P@10	0.3105	0.3433	0.3582	0.3507(-2.1)
	NL-ES	MAP	0.3578	0.3722	0.369	0.3875(4.1) [†]
		P@5	0.3821	0.406	0.4119	0.406(0.0)
		P@10	0.3105	0.3478	0.3448	0.3537(1.7)
NL-FR	MAP	0.3578	0.3722	0.3553	0.3875(4.1) [†]	
	P@5	0.3821	0.406	0.397	0.409(0.7)	
	P@10	0.3105	0.3478	0.3433	0.3463(-0.4)	
ES-FR	MAP	0.3578	0.369	0.3553	0.3823(3.6)	
	P@5	0.3821	0.4119	0.397	0.4119(0.0)	
	P@10	0.3105	0.3448	0.3433	0.3418(-0.9)	
FR	EN-ES	MAP	0.4356	0.4658	0.4634	0.4803(3.1)
		P@5	0.4776	0.4925	0.4925	0.4985(1.2)
		P@10	0.4194	0.4358	0.4388	0.4493(3.1) [†]

Table 3: Comparison of MultiPRF Multiple Assisting Languages using parallel assistance framework with MultiPRF with single assisting language. Only language pairs where positive improvements were obtained are reported here. Results marked as [†] indicate that the improvement was statistically significant over baseline (Maximum of MultiPRF with single assisting language) at 90% confidence level ($\alpha = 0.01$) when tested using a paired two-tailed t-test.

uniformly choose the top ten documents for feedback.

6 Results and Discussion

Tables ?? and ?? present the results for MultiPRF with two assisting languages using parallel assistance and selective assistance framework. Out of the total 60 possible combinations, in Table ??, we only report the combinations where we have obtained positive improvements greater than 3%. We observe most improvements in English, Finnish and French. We did not observe any improvements using the serial assistance framework over MultiPRF with single assisting lan-

Source Langs	Assist. Langs		Parallel Model	Selective Model
EN	DE-NL	MAP	0.4651	0.4848
		P@5	0.5254	0.5224
		P@10	0.4493	0.4522
	NL-FI	MAP	0.4387	0.4502
		P@5	0.5015	0.5164
		P@10	0.4284	0.4358
DE	EN-FR	MAP	0.4097	0.4302
		P@5	0.594	0.5851
		P@10	0.5149	0.5179
	FR-ES	MAP	0.4215	0.4333
		P@5	0.591	0.591
		P@10	0.5239	0.5209
	FR-NL	MAP	0.4139	0.4236
		P@5	0.5701	0.5701
		P@10	0.5075	0.5134
	FR-FI	MAP	0.3925	0.4055
		P@5	0.5101	0.5642
		P@10	0.4851	0.5
	NL-FI	MAP	0.3974	0.4192
		P@5	0.5731	0.5612
		P@10	0.497	0.503
ES	EN-FI	MAP	0.4436	0.4501
		P@5	0.6179	0.6269
		P@10	0.5567	0.5657
	DE-FI	MAP	0.4542	0.465
		P@5	0.6269	0.6179
		P@10	0.5627	0.5582
	NL-FI	MAP	0.4531	0.4611
		P@5	0.6269	0.6299
		P@10	0.5627	0.5627

Table 4: Results showing the positive improvements of MultiPRF with selective assistance framework over MultiPRF with parallel assistance framework.

guage. Hence, we do not report their results as the results were almost equivalent to single assisting language. As shown in Table ??, selective assistance does give decent improvements in some language pairs. An interesting point to note in selective assistance is that it helps languages like Spanish whose monolingual performance and document coverage are both high.

6.1 Qualitative Comparison of Feedback Terms using Multiple Languages

In this section, we qualitatively compare the results of MultiPRF with two assisting languages with that of MultiPRF with single assisting language, based on the top feedback terms obtained by each model. Specifically, in Table 5 we compare the terms obtained by MultiPRF using (i) Only L_1 as assisting language, (ii) Only L_2 as assisting language and (iii) Both L_1 and L_2 as assisting languages in a parallel combination. For example, the first row in the above table shows the terms obtained by each model for the English query “Golden Globes 1994”. Here, L_1 is French and L_2 is Spanish. Terms like “Gold” and “Prize” appearing in the translated feedback model of L_1 cause a drift in the topic towards

“Gold Prize” resulting in a lower MAP score (0.33). Similarly, the terms like “forrest” and “spielberg” appearing in the translated feedback model of L_2 cause a drift in topic towards *Forrest Gump and Spielberg Oscars* resulting in a MAP score (0.5). However, when the models from two languages are combined, terms which cause a topic drift get ranked lower and as a result the focus of the query is wrenched back. A similar observation was made for the English query “Damages in Ozone Layer” using French (L_1) and Spanish (L_2) as assisting languages. Here, terms from the translated feedback model of L_1 cause a drift in topic towards “militri bacteria” whereas the terms from the translated feedback model of L_2 cause a drift in topic towards “iraq war”. However, in the combined model these terms get lower rank there by bringing back the focus of the query. For the Finnish query “Lasten oikeudet” (Children’s Rights), in German (L_1), the topic drift is introduced by terms like “las, gram, yhteis”. In case of Dutch (L_2), the query drift is caused by “mandy, richard, slovakia” (L_2) and in the case of combined model, these terms get less weightage and the relevant terms like “laps, oikeuks, vanhemp” which are common in both models, receive higher weightage causing an improvement in query performance.

Next, we look at a few negative examples where the parallel combination actually performs poorer than the individual models. This happens when some *drift-terms* (i.e., terms which can cause topic drift) get mutually reinforced by both the models. For example, for the German query “Konkurs der Baring-Bank” (Bankruptcy of Baring Bank) the term “share market” which was actually ranked lower in the individual models gets boosted in the combined model resulting in a drift in topic. Similarly, for the German query “Ehren-Oscar für italienische Regisseure” (Honorary Oscar for Italian directors) the term “head office” which was actually ranked lower in the individual models gets ranked higher in the combined model due to mutual reinforcement resulting in a topic drift.

TOPIC NO.	QUERIES (Meaning in Eng.)	TRANSLATED ENGLISH QUERIES (Assisting Lang.)	L1 MAP	L2 MAP	L1-L2 MAP	Representative Terms with L1 as Single Assisting Language (With Meaning)	Representative Terms with L2 as Single Assisting Language (With Meaning)	Representative Terms with L1& L2 as Assisting Langs. (With Meaning)
English '03 TOPIC 165	Globes 1994	Golden Globes 1994 (FR) Globos de Oro 1994 (ES)	0.33	0.5	1	Gold, prize, oscar, nomin, best award, hollywood, actor, director ,actress, world, won ,list, winner, televi, foreign ,year, press	world, nomin, film, award, delici, planet, earth, actress, list, drama, director, actor, spielberg, music, movie, Forrest, hank	oscar, nomin, best, award, hollywood actor, director, cinema, televi, music, actress, drama, role, hank, foreign, gold
Finnish '03 TOPIC 152	Lasten oikeudet (Children's Rights)	Rechte des Kindes (DE) Kinderrechten (NL)	0.2	0.25	0.37	laps (child), oikeuks (rights), oikeud (rights), kind, oikeus (right), isä (father), oikeut (justify), vanhemp (parent), vanhem (parents), las, gram, yhteis, unicef, sunt, äiti(mother), yleissopimnks(conventions)	oikeuks (rights), laps (child), oikeud (right), mandy, richard, slovakia, tähänast (to date), tuomar (judge), tyto, kid, , nuor (young people), nuort (young), sano(said) , perustam(establishing)	laps (child), oikeuks (rights), oikeud (rights), oikeus (right), isä (father, parent), vanhemp (parent), vanhem (parents), oikeut (justify), las, mandy, nuort (young), richard, nuor (young people), slovakia, tähänast (to date),
English '03 TOPIC 148	Damages in Ozone Layer	Dommages à la couche d'ozone (FR) Destrucción de la capa de ozono (ES)	0.08	0.07	0.2	damag, militri, uv, layer, condition, chemi, bacteria, ban, radiat, ultraviolet	damag, weather, atmospher, earth, problem, report, research, harm, iraq, war, scandal, illigel, latin, hair	damag, uv, layer,weather, atmospher, earth, problem, report, research , ultraviolet, chemi
German '03 TOPIC 180	Konkurs der Baring-Bank (Bankruptcy of Baring Bank)	Bankruptcy of Barings (EN) Baringsin Konkurssi (FI)	0.55	0.51	0.33	zentralbank(central bank),bankrott(bankruptcy), investitionsbank, sigapur, london , britisch, index, tokio, england, werbung(advertising), japan	fall, konkurs, bankrott(Bankruptcy), warnsignal(warning), ignoriert, zusammenbruch(collepse), london, singapur, britisch(british), dollar, tokio, druck(pressur), handell(trade)	aktienmarkt(share market), investitionsbank, bankrott, zentralbank(central bank), federal, singapur, london, britisch, index, tokio, dollar, druck, england, dokument(document)
German '03 TOPIC 198	Ehren-Oscar für italienische Regisseure (Honorary Oscar for Italian directors)	Honorary Oscar for Italian Directors (EN) Kunnia-Oscar italialaisille elokuvaohjaajille (FI)	0.5	0.35	0.2	Direktor(director), film, regierungscheff(prime), best antonionis, antonionins, lieb, geschicht(history), paris, preis, berlin, monitor, kamera	Generaldirektion(General director), film, ehrenmitglied, regisseur, direktor, verleih , itali, oscar, award, antonionins	generaldirektion(head office), ehrenmitglied(honorable member), regierungschef(prime), regisseur(director),oscar, genossenschaftsbank (corporate bank)

Table 5: Qualitative Comparison of MultiPRF Results using two assisting languages with single assisting language.

6.2 Effect of Coverage on MultiPRF Accuracy

A study of the results obtained for MultiPRF using single assisting language and multiple assisting languages with different source languages showed that certain languages are more suited to be benefited by assisting languages. In particular, languages having smaller collections are more likely to be benefited if assisted by a language having a larger collection size. For example, Finnish which has the smallest collection (55344 documents) showed maximum improvement when supported by assisting language(s). Based on this observation, we plotted a graph of the collection size of a source language v/s the average improvement obtained by using two assisting languages to see if there exists a correlation between these two factors. As shown in Figure 3, there indeed exists a high correlation between these two entities. At one extreme, we have a language like Spanish which has the largest collection (454045 documents) and is not benefited much by assisting languages. On the other extreme, we have Finnish which has the smallest collection size and is benefited most by assisting languages.

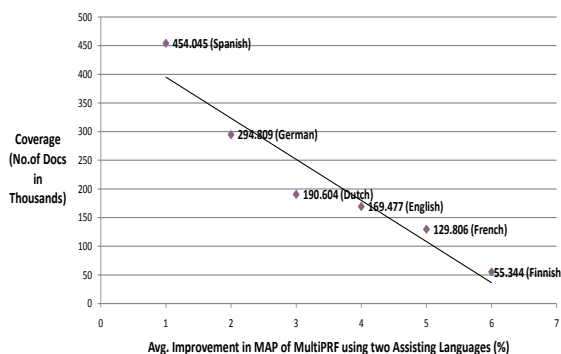


Figure 3: Effect of Coverage on Average MultiPRF MAP using Two Assisting Languages.

6.3 Effect of Number of Assisting Languages on MultiPRF Accuracy

Another interesting question which needs to be addressed is “Whether it helps to use more than two assisting languages?” and if so “Is there an optimum number of assisting languages beyond which there will be no improvement?”. To answer these questions, we performed experiments using 1-4 assisting languages with each source language. As seen in Figure 4, in general as the number of assisting languages increases the performance saturates (typically after 3 languages). Thus, for 5 out of the 6 source languages, the performance saturates after 3 languages which is in line with what we would intuitively expect. However, in the case of German, on an average, the

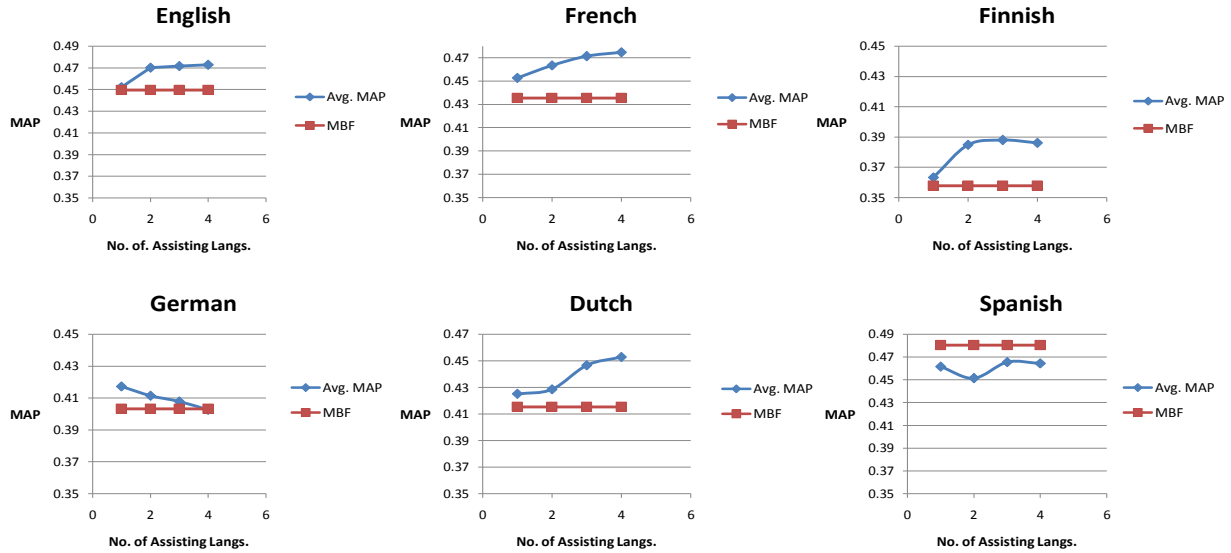


Figure 4: Effect of Number of Assisting Languages on Avg. MultiPRF Performance with Multiple Assistance.

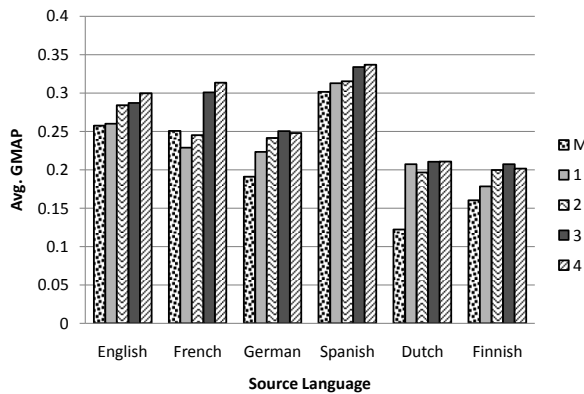


Figure 5: Effect of Number of Assisting Languages on Robustness measured through GMAP.

performance drops as the number of assisting languages is increased. This drop is counter intuitive and needs further investigation.

6.4 Effect of Number of Assisting Languages on Robustness

One of the primary motivations for including multiple assisting languages in MultiPRF was to increase the robustness of retrieval through better coverage. We varied the number of assisting languages for each source and studied the average GMAP. The results are shown in Figure 5. We observe that in almost all the source languages, the GMAP value increases with number of assisting languages and then reaches a saturation after reaching three languages.

7 Conclusion

In this paper, we extended the MultiPRF framework to multiple assisting languages. We presented three different configurations for including multiple assisting languages - a) Parallel b) Serial and c) Selective. We observe that the results are mixed with parallel and selective assistance showing improvements in some cases. We also observe that the robustness of MultiPRF increases with number of assisting languages. We analyzed the influence of coverage of MultiPRF accuracy and observed that it is inversely correlated. Finally, increasing the number of assisting languages increases the MultiPRF accuracy to some extent and then it saturates beyond that limit. Many of the above results (negative results of serial, selective configurations etc.) require deeper investigation which we plan to take up in future.

References

- Braschler, Martin and Carol Peters. 2004. Cross-language evaluation forum: Objectives, results, achievements. *Inf. Retr.*, 7(1-2):7-31.
- Buckley, Chris, Gerald Salton, James Allan, and Amit Singhal. 1994. Automatic query expansion using smart : Trec 3. In *Proceedings of The Third Text REtrieval Conference (TREC-3)*, pages 69-80.
- Chinnakotla, Manoj K., Karthik Raman, and Pushpak Bhattacharyya. 2010a. Multilingual pseudo-

- relevance feedback: English lends a helping hand. In *ACM SIGIR 2010*, Geneva, Switzerland, July. ACM.
- Chinnakotla, Manoj K., Karthik Raman, and Pushpak Bhattacharyya. 2010b. Multilingual pseudo-relevance feedback: Performance study of assisting languages. In *ACL 2010*, Uppsala, Sweden, July. ACL.
- Dempster, A., N. Laird, and D. Rubin. 1977. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society*, 39:1–38.
- Hawking, David, Paul Thistlewaite, and Donna Harman. 1999. Scaling up the trec collection. *Inf. Retr.*, 1(1-2):115–137.
- John Lafferty and Chengxiang Zhai. 2003. Probabilistic Relevance Models Based on Document and Query Generation. In *Language Modeling for Information Retrieval*, volume 13, pages 1–10. Kluwer International Series on IR.
- Mitra, Mandar, Amit Singhal, and Chris Buckley. 1998. Improving automatic query expansion. In *SIGIR '98: Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 206–214, New York, NY, USA. ACM.
- Och, Franz Josef and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1):19–51.
- Ounis, I., G. Amati, Plachouras V., B. He, C. Macdonald, and Johnson. 2005. Terrier Information Retrieval Platform. In *Proceedings of the 27th European Conference on IR Research (ECIR 2005)*, volume 3408 of *Lecture Notes in Computer Science*, pages 517–519. Springer.
- Philipp, Koehn. 2005. Europarl: A parallel corpus for statistical machine translation. In *MT Summit*.
- Robertson, Stephen. 2006. On gmap: and other transformations. In *CIKM '06: Proceedings of the 15th ACM international conference on Information and knowledge management*, pages 78–83, New York, NY, USA. ACM.
- Voorhees, Ellen. 2006. Overview of the trec 2005 robust retrieval track. In *E. M. Voorhees and L. P. Buckland, editors, The Fourteenth Text REtrieval Conference, TREC 2005*, Gaithersburg, MD. NIST.
- Wu, Dan, Daqing He, Heng Ji, and Ralph Grishman. 2008. A study of using an out-of-box commercial mt system for query translation in clir. In *iNEWS '08: Proceeding of the 2nd ACM workshop on Improving non english web searching*, pages 71–76, New York, NY, USA. ACM.
- Xu, Jinxi and W. Bruce Croft. 2000. Improving the effectiveness of information retrieval with local context analysis. *ACM Trans. Inf. Syst.*, 18(1):79–112.
- Zhai, Chengxiang and John Lafferty. 2001. Model-based Feedback in the Language Modeling approach to Information Retrieval. In *CIKM '01: Proceedings of the tenth international conference on Information and knowledge management*, pages 403–410, New York, NY, USA. ACM Press.
- Zhai, Chengxiang and John Lafferty. 2004. A Study of Smoothing Methods for Language Models applied to Information Retrieval. *ACM Transactions on Information Systems*, 22(2):179–214.