Positional Translation Language Model for Ad-Hoc Information Retrieval

Xinhui Tu[1,2], Jing Luo[1,2], Bo Li[3], Tingting He[3], and Jinguang Gu[1,4]

[1] College of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan, China
[2] Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan, China
[3] Department of Computer Science, Central China Normal University, Wuhan, China
[4] State Key Lab of Software Engineering, Wuhan University, Wuhan, China
{tuxinhui,luoluocat,simongu}@gmail.com, liboccnu@126.com, tthe@mail.ccnu.edu.cn

Abstract. Most existing language modeling approaches are based on the term independence hypothesis. To go beyond this assumption, two main directions were investigated. The first one considers the use of the proximity features that capture the degree to which search terms appear close to each other in a document. Another one considers the use of semantic relationships between words. Previous studies have proven that these two types of information, including term proximity features and semantic relationships between words, are both useful to improve retrieval performance. Intuitionally, we can use them in combination to further improve retrieval performance. Based on this idea, this paper propose a positional translation language model to explicitly incorporate both of these two types of information under language modeling framework in a unified way. In the first step, we present a proximity-based method to estimate word-word translation probabilities. Then, we define a translation document model for each position of a document and use these document models to score the document. Experimental results on standard TREC collections show that the proposed model achieves significant improvements over the state-of-the-art models, including positional language model, and translation language models.

Keywords: Positional Language Model, Translation Language Model, Information Retrieval.

1 Introduction

Language modeling (LM) for Information Retrieval (IR) has been a promising area of research over the past decade and a half. It provides an elegant mathematical model for ad-hoc text retrieval with excellent empirical results reported in the literature [12][20]. However, language models suffer from one problem: term independence assumption which is common for all retrieval models.

To address this problem, two main directions were investigated. The first one is based on the use of the proximity features. These features capture the degree to which

V.S. Tseng et al. (Eds.): PAKDD 2014, Part II, LNAI 8444, pp. 134–145, 2014.

search terms appear close to each other in a document. To incorporate the cues of term position and term proximity under language model framework, Lv and Zhai [10] proposed a positional language model (PLM). In PLM, a language model for each term position in a document is defined, and document is scored based on the scores of its PLMs.

The second one considers the use of semantic relationships between words. In order to reduce the semantic gap between documents and queries, statistical translation models (TLM) have been proposed for information retrieval to capture semantic word relations [2]. The basic idea of translation language models is to estimate the probabilities of translating a word in a document to query words. Since a word in a document could be translated into different words in the query, translation language models can avoid exact matching of words between documents and queries.

The previous studies have proven that term proximity features and semantic relationships between words are both useful information to improve retrieval performance (e.g., [2][7][10][11]). Intuitionally, we can use them in combination to further improve retrieval performance. Based on this idea, this paper proposes a positional translation language model to explicitly incorporate these two types of information in a united way. In the first step, we present a proximity-based method to estimate word-word translation probabilities. Then, we define a translation document model for each position of a document and use these document models to score the document.

The main contribution of this paper is as follows： First, we propose a new proximity-based method, in which the proximity of co-occurrences is taking into account, to estimate word-word translation probabilities. Second, we propose a positional translation language model (PTLM) to explicitly incorporate term proximity features and semantic relationships between words in a unified way. Finally, extensive experiments on standard TREC collections have been conducted to evaluate the proposed model. Experimental results on standard TREC collections show that PTLM achieves significant improvements over the state-of-the-art models, including positional language model, and translation language models.

2 Background: PLM and TLM

2.1 Basic Language Modelling Approach

The basic idea of language models is to view each document to have its own language model and model querying as a generative process. Documents are ranked based on the probability of their language model generating the given query. Different implementations were proposed [20]. The general ranking formula is defined as follows:

$$logp(D|Q) \stackrel{\text{rank}}{=} \sum_{w \in V} c(w,Q)\, logp(w|D) \tag{1}$$

where $\stackrel{\text{rank}}{=}$ means equivalence for the purpose of ranking documents, $c(w,Q)$ is the count of word w in query Q, and V is the vocabulary set. The challenging part is to estimate a document model $p(w|D)$. The simplest way to estimate $p(w|D)$ is the maximum likelihood estimator. However, this method is suffering from the data sparseness problem. To address this problem, some effective smoothing approaches, which combine the document model with the background collection model, have been proposed. One commonly used method is Dirichlet Prior smoothing methods [18], which is defined as follows:

$$p(w|Q) = \frac{|D|}{|D| + \mu} p_{ml}(w|D) + \frac{\mu}{|D| + \mu} p_{ml}(w|C) \tag{2}$$

2.2 Positional Language Model

To incorporate the cues of term position and term proximity under language model framework, Lv and Zhai [10] proposed a positional language model (PLM). In PLM model, for each document $D(w_1; \ldots; w_i; \ldots; w_j; \ldots; w_N)$, where *1, i, j,* and *N* are absolute positions of the corresponding terms in the document, and *N* is the length of the document, a virtual $\mathrm{D_i}$ document is estimated at each position. This model is represented as a term frequency vector $D\langle c'(w_1; i); \ldots; c'(w_N; i)\rangle$, where $c'(w; i)$ is the total propagated count of term w at position i from the occurrences of w in all the positions. That is $c'(w, i) = \sum_{j=1}^{N} c(w, j) k(i, j)$ where $\mathrm{c}(w, j)$ is the count of term w at position i in document *D*. If w occurs at position i, it is 1, otherwise 0. $k(i, j)$ is the propagated count to position i from a term at position j. Several proximity-based density functions are used to estimate this factor: (Gaussian kernel, Triangle kernel, Circle kernel, Cosine kernel). Once the virtual document D_i is estimated, the language model of this virtual document can be estimated as follow

$$p(w|D, i) = \frac{c'(w, i)}{\sum_{w' \in V} c'(w', i)} \tag{3}$$

where *V* is the vocabulary, $p(w|D, i)$ is noted as a positional language model at position i. To compute the final score of document D, they used the position-specific scores. Different strategies were used: Best Position Strategy, Multi-Position Strategy, Multi-σ Strategy.

2.3 Statistical Translation Language Model

To incorporate the semantic relationship between terms under language model framework, Berger and Lafferty proposed translation language modelling approach to estimate $p(w|D)$ based on statistical machine translation [2]. In this approach, the document model $p(w|D)$ can be calculated by using the following "translation document model":

$$p_t(w|D) = \sum_{u \in D} p_t(w|u)\, p(u|D) \tag{4}$$

where $p(u|D)$ is the probability of seeing word u in document d, and $p_t(w|u)$ is the probability of "translating" word u into word w. In this way, a word can be translated into its semantically related words with non-zero probability, which allows us to score a document by counting the matches between a query word and semantically related words in document.

The key part for translation language model is estimating translation probabilities. Berger and Lafferty [2] proposed a method to estimate translation probabilities by generating synthetic query. This method is inefficient and does not have good coverage of query words. In order to overcome these limitations, Karimzadehgan and Zhai [7] proposed an effective estimation method based on mutual information. Recently, Karimzadehgan and Zhai [8] defined four constraints that a reasonable translation language model should satisfy, and proposed a new estimation method which is shown to be able to better satisfy the constraints. This new estimation method, namely conditional context analysis, is described in formula 5.

3 Positional Translation Language Model

In this section, we will describe the PTLM in detail. In the first part, a proximity-based method is presented to estimate word-word translation probabilities. Then, we will introduce how to estimate the translation document model for each position within a document. Finally, these positional document models are used to score the document.

3.1 Estimating Translation Probability

In the conditional context analysis method proposed in [8], the probability of translating word u into word w can be estimated as follows:

$$p(w|u) = \frac{c(w,u) + 1}{\sum_{w'} c(w',u) + |V|} \tag{5}$$

where $c(w,u)$ is the co-occurrences of word u with word w, and $|V|$ is the size of the vocabulary.

In this method, any co-occurrence within the document is treated in the same way, no matter how far they are from each other. This strategy is not optimal as a document may cover several different topics and thus contain much irrelevant information. Intuitionally, closer words usually have stronger relationships, thus should be more relevant. Therefore, we introduce a new concept, namely proximity-based word co-occurrence frequency (pcf) to model the proximity feature of co-occurrences.

Recently, density functions based on proximity are proven to be effective to characterize term influence propagation. A number of term propagation functions (e.g. Gaussian, Triangle, Cosine and Circle) have been proposed [10][11]. In this section, we adopted Gaussian functions because it has been shown to be effective in most cases. The Gaussian-based pcf can be calculated as follows:

$$pcf(w,u) = \sum_{D \in Col(w,u)} exp\left[\frac{-(dist(w,u,D))^2}{2\sigma^2}\right] \tag{6}$$

where, σ is a parameter in Gaussian distribution, $Col(w,u)$ is the set of documents which contain both w and u, and $dist(w,u,D)$ is the distance score of word w and u in document D.

In this paper, three commonly used distance measures are adopted to late $dist(w,u,D)$. We will use the following short document D as an example to explain how to calculate distance score in the three distance measures.

	1	2	3	4	5	6	7	8	9	10	
D = {	w	u	c	k	w	u	k	e	w	g	}

Minimum pair distance: It is defined as the minimum distance between any occurrences of w and u in document D. In the example, $dist(w,u,D)$ is 1 and can be calculated from the position vectors.

Average pair distance: It is defined as the average distance between w and u for all position combinations in D. In the example, the distances from the first occurrence of w (in position 1) to all occurrences of u are: $\{1$ and $5\}$. This is computed for the next occurrence of w (in position 5) and so on. $dist(w,u,D)$ for the example is (((2-1) + (6-1)) + ((5-2) + (6-5)) + ((9-2) + (9-6)))/(2 · 3) = 20/6 = 3.33.

Average minimum pair distance: It is defined as the average of the shortest distance between each occurrence of the least frequently occurring word and any occurrence of the other word. In the example, u is the least frequently occurring word so $dist(w,u,D)$ = ((2–1)+(6–5))/2 =1.

Then, the probability of translating word u into word w can be estimated as follows:

$$p(w|u) = \frac{pcf(w,u) + \epsilon}{\sum_{w'} pcf(w',u) + |V| * \epsilon} \tag{7}$$

where ϵ is a smoothing parameter in order to account for unseen words in the context of u. Here ϵ is set equals to the smallest of all pcf values in collection.

In order to satisfy the constraints defined in [8], we adjust self-translation probabilities as follows:

$$p_t(u|u) = s \ (s \geq 0.5) \tag{8}$$

$$p_t(w|u) = (1-s) * \frac{p(w|u)}{\sum_{v \neq u} p(v|u)} \tag{9}$$

where parameter s is a constant value that could be set to 0.5 <= s <= 1. Note that when s = 1, the query likelihood model are gained.

3.2 Estimating Translation Document Model

The state-of-art translation language models use an entire document as a unit to estimate the generative probability of the query [7][8]. This strategy is not optimal as a document may cover several different topics. Intuitionally, the words referring to the same topic may occur close to each other. Positional language model has been proven to be an effective way to incorporate the cues of term position and term proximity under language model framework [10]. In this section, we will introduce a positional translation language model to naturally incorporate two types of information, including term proximity features and semantic relationships between terms, under language model framework in a united way.

The key idea of our method is to extend the translation language model from document level to positional level via the positional language model. The proposed model can capture the topic of the document at the position by giving more weight on words close to the position and less weight on words far away. The translation language model at each position can be estimated based on all the propagated counts of all the words to the position as if all the words had appeared actually at the position with discounted counts.

Previous studies have shown that translation language model works better with Dirichlet prior smoothing [7][8]. Therefore, in the rest of the paper, we further focus on PTLM with Dirichlet prior smoothing only. The final positional translation language model for position i in document D can be defined as follows:

$$p_t(w|D,i) = \frac{|D|}{|D|+\mu}\left[\sum_{u \epsilon D} p_t(w|u)\, p(u|D,i)\right] + \frac{\mu}{|D|+\mu} p(w|C) \tag{10}$$

where $p_t(w|u)$ is the translation probability from word u to word w, and can be estimated by formula 8 and 9; $p(u|D,i)$ is the positional document model at position i of document D, and can be estimated as follows:

$$p(u|D,i) = \frac{c'(u,i)}{\sum_{u' \in V} c'(u',i)} \tag{11}$$

where $c'(u,i)$ is the total propagated count of term u at position i from the occurrences of u in all the positions. $c'(u,i)$ can be estimated using the Gaussian kernel function:

$$c'(u,i) = \sum_{j=1}^{|D|} c(u,j) \exp\left[\frac{-(i-j)^2}{2\sigma^2}\right] \tag{12}$$

where i and j are absolute positions of the corresponding terms in document, and $|D|$ is the length of the document, $c(u,j)$ is the real count of term u at position j.

3.3 Ranking Document

In the section 3.2, we have obtained a translation language model for each position in a document. Intuitively, we can imagine that the PTLMs give us multiple representations of D. Thus given a query Q, we can adopt the KL-divergence retrieval model [19] to score each PTLM as follows:

$$S(Q,D,i) = -\sum_{w \in V} p(w|Q) log \frac{p(w|Q)}{p_t(w|D,i)} \quad (13)$$

Then, the position-specific scores can be used to compute the final score of document D. In this paper, we compute the final score of document D using the best position strategy [10], which simply scores a document based on its best match position and can be defined as follows:

$$S(Q,D) = max_{i \in [1,N]}\{S(Q,D,i)\} \quad (14)$$

4 Experiments

4.1 Data Set

We used six standard TREC data sets in our study. They represent different sizes and genre of text collections. Table 1 shows some basic statistics about these data sets. Each document is processed in a standard way for indexing. Words are stemmed (using porter-stemmer), and stop words are removed. In the experiments, we only use title of the queries because semantic word matching is necessary for such short queries.

Table 1. Document set characteristic

	TREC7	DOE	WSJ	TREC8	AP88-89	FR
queries	351-400	51-100	51-100	401-450	51-100	51-100
#doc	528,155	226,087	74,520	528,155	164,597	45,820

In each experiment, we use the KL-divergence model using Dirichlet prior smoothing (with prior parameter μ=1000) to retrieve 2000 documents for each query, and then use the PTLM to re-rank them. The top-ranked 1000 documents are used for comparison with other models. In order to evaluate our model and compare it to other models we use the MAP measure, which is widely accepted measure for evaluating effectiveness of ranked retrieval systems.

In the section 3.1, three different proximity measures are adapted to measure the distance score of two words in a document. The corresponding models based on the three different proximity measures are evaluated on standard TREC collections. The methods used for the experiments are:

- **QL:** baseline, query likelihood model with Dirichlet prior smoothing [18].
- **KL:** baseline, KL-divergence model with Dirichlet prior smoothing [19].
- **TM-MI:** translation language model with mutual information [7].
- **TM-CCON:** translation language model with conditional context analysis [8].
- **PLM:** positional language model with the best position strategy [10].
- **PTLM-1:** PTLM with minimum pair distance.

- **PTLM-2:** PTLM with average pair distance.
- **PTLM-3:** PTLM with average minimum pair distance.

4.2 Comparing with Existing Retrieval Models

As we can see from all the PTLM models used in our experiments, there are several controlling parameters to tune. In order to make the comparison fair, we evaluate PTLMs and PLM by a 5-fold cross-validation on each collection. For the two baselines (QL and KL), parameter μ in the Dirichlet smoothing is set to the optimal value for each collection. The results of TM-MI, TM-CCON are directly from [8].

Table 2 shows the results for these models with Dirichlet prior smoothing. Comparing the rows in the table indicates that the PTLM models achieve significant improvements over the state-of-the-art models, including positional language model and translation language models. In addition, the results confirm our hypothesis that the two types of information can be used in combination to improve retrieval performance. Comparing the three variants of PTLM, PTLM-3 is more effective and robust than PTLM-2 and PTLM-1. It also indicates that average-minimum-pair distance measure can capture the proximity feature of co-occurrences better than the other two measures. The significance test results using Wilcoxon signed-rank test indicate that the differences between the PTLM models and the start-of-art models are statistically significant.

Table 2. The comparison of experiment results
(* and $^+$ mean improvements over TM-CCON and PLM are statistically significant with Wilcoxon signed-rank test, respectively)

	TREC7	DOE	WSJ	TREC8	AP88-89	FR
QL	0.1852	0.1740	0.2600	0.2518	0.2154	0.2817
KL	0.1847	0.1742	0.2584	0.2509	0.2196	0.2697
TM-MI	0.1854	0.1750	0.2658	-	-	-
TM-CON	0.1920	0.1844	0.2780	-	-	-
PLM	0.1893	0.1795	0.2641	0.2548	0.2196	0.2842
PTLM-1	0.2003^{*+}	0.1952^{*+}	0.2896^{*+}	0.2672^{+}	0.2246^{+}	0.2885^{+}
PTLM-2	0.2021^{*+}	0.1967^{*+}	0.2913^{*+}	0.2685^{+}	0.2259^{+}	0.2891^{+}
PTLM-3	$\mathbf{0.2030}^{*+}$	$\mathbf{0.1975}^{*+}$	$\mathbf{0.2924}^{*+}$	$\mathbf{0.2692}^{+}$	$\mathbf{0.2276}^{+}$	$\mathbf{0.2920}^{+}$

4.3 Parameter Sensitivity Study

An important issue that may affect the robustness of the PTLM models is the sensitivity of their parameters s (in Equation 8, 9) and σ (in Equation 6, 12). The parameter s controls the amount of self-translation probabilities. The kernel parameter σ in

Equation 6 determines the distance in which words are considered to be related. Another kernel parameter σ in Equation 12 restricts the propagation scope of a virtual document. In this section, we study how sensitive these parameters are to MAP measure.

We investigate a large range of σ (in Equation 6) from 10 to 1000. Generally, the value of σ affects the performance of all PTLM models extensively. The experimental results show that the influence of σ is collection-based. For the three PTLM models, their curves fluctuate similarly on the same collection. However, the best σ values for these PTLM models are not the same. For example, on the TREC7 collection, optimal σ value for PTLM-1 is 80, and the corresponding value for PTLM-2 is 150. Thus, the optimal values of σ depend on the proximity measures and the collections. Figure 1 plots the evaluation metrics MAP obtained by the three PTLM models with σ values ranging from 10 to 1000 on TREC7.

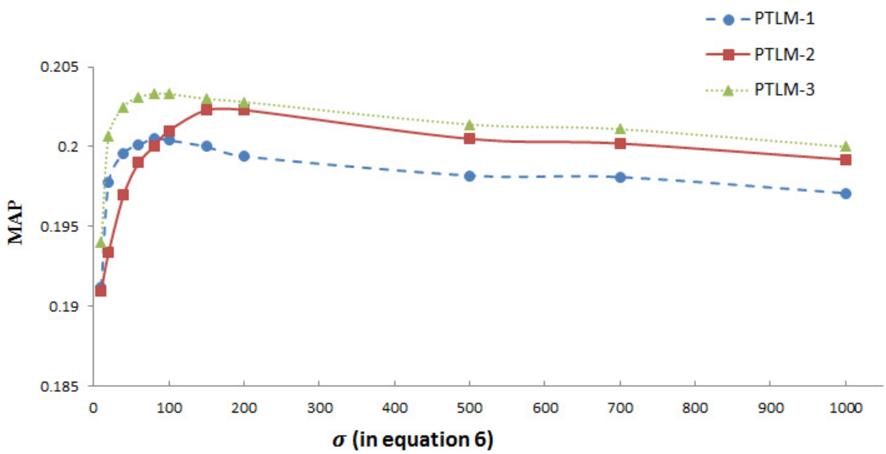

Fig. 1. PTLM-1, PTLM-2, PTLM-3 over TREC7 with **σ** (in Equation 6) values ranging from10 to 1000

The experimental results also show that the influence of s is collection-based. For one collection, the best s values for all PTLM models are the same. Specifically, the best s values are 0.7, 0.8, and 0.5 for TREC7, DOE, and WSJ, respectively.

In order to see how the propagation scope parameter σ (in Equation 12) affects the performance of the PTLM models, we test a set of values from 25 to 275 in increments of 25. Overall, we see that a relatively large often brings the best performance. It also seems that the performance of the PTLM models stabilizes after σ reach 175.

To investigate how the Dirichlet prior parameter μ (in Equation 10) affects the performance of the PTLM models, we also change the settings of the smoothing parameters for them. The results indicate that the optimal smoothing parameters are the same (equals to 500) for all the three PTLM models on all collections.

5 Related Work

Most existing information retrieval model including probabilistic and vector space models are based on the term independence hypothesis. Given common knowledge about language, such an assumption might seem unrealistic. To go beyond the term independency assumption in information retrieval, two main directions were investigated.

The first one considers the use of the proximity features that capture the degree to which search terms appear close to each other in a document. For example, it looks at the minimum span of the query terms appearing in the document. Term proximity, as an effective retrieval heuristic, has been studied extensively in the past few years. In these papers, various methods have been proposed to integrate proximity information into different retrieval models. Keen [9] firstly attempted to import term proximity in the Boolean retrieval model by introducing a "NEAR" operator. Buttcher et al. [3] proposed an integration of term proximity scoring into Okapi BM25 and obtain improvements on several collections. Tao et al. [14] systematically studied five proximity measures and compared their performance in various retrieval models. Zhao et al. [16] used a query term's proximity centrality as a hyper parameter in Dirichlet language model under the language modelling framework. Lv and Zhai [10] integrated the position and proximity information into the language model by defining a language model for each position within a document. Zhao et al. [17] introduce a pseudo term, namely cross term, to model term proximity for boosting retrieval model. Miao et al. [11] has attempted to incorporate proximity information into the Rocchio's model.

The second one considers the use of semantic relationships between words. Under this way, relevant words are used to enrich document or query representation. Many studies have tried to bridge the vocabulary gap between documents and queries both based on term co-occurrences [1, 6, 13] and hand-crafted thesaurus [15]. Some other works have considered to combine both approaches [4]. Berger and Lafferty [2] firstly proposed a translation language model to corporate semantic relationship between words under the language modeling framework. To train translation models, they used synthetically generated query-document pairs. An alternation way of estimating the translation model is based on document titles [5]. Recent works have relied on document-based word co-occurrences to estimate the translation model [7][8].

6 Conclusion

Term proximity features and semantic relationships between words have proven to be two kinds of useful information to improve retrieval performance. In this paper, we proposed a positional translation language model to incorporate both of them in a unified way. In the first step, a new proximity-based method is presented to estimate the translation model. Three proximity measures are then adopted for calculating the distance score of two words within a document. The corresponding models based on these measures, PTLM-1, PTLM-2 and PTLM-3, are evaluated on six standard TREC collections. Our experiment results indicate that the PTLM models are more effective than the state-of-art models, including positional language model and translation language models. Comparing the three variants of PTLM, PTLM-3 is more effective than the other two.

Since the number of positions is much larger than the number of documents, the cost of estimating PTLMs can be extremely high. For the sake of efficiency, we use PTLM to re-rank the top 2000 documents from initial search results. However, such a strategy does not fully take advantage of the capacity of PTLM to potentially retrieve relevant documents that do not match any query word. In the future, we will try to study how to reduce the computational complexity of PTLM and to further improve retrieval performance.

Acknowledgments. This work was partially supported by the National Science Foundation of China under grants No. 60803160 and No. 61300144, the Key Projects of National Social Science Foundation of China under grant number 11ZD&189. It was partially supported by NSF of Hubei Prov. under grant number 2013CFB334, and the State Key Lab of Software Engineering Open Foundation of Wuhan University under grant number SKLSE2012-09-07.

References

1. Bai, J., Song, D., Bruza, P., Nie, J.Y., Cao, G.: Query expansion using term relationships in language models for information retrieval. In: Proceedings of the 14th ACM International Conference on Information and Knowledge Management, pp. 688–695. ACM, New York (2005)
2. Berger, A., Lafferty, J.: Information retrieval as statistical translation. In: Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 222–229. ACM, New York (1999)
3. Büttcher, S., Clarke, C.L.A., Lushman, B.: Term proximity scoring for ad-hoc retrieval on very large text collections. In: Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 621–622. ACM, New York (2006)
4. Cao, G., Nie, J.Y., Bai, J.: Integrating word relationships into language models. In: Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 298–305. ACM, New York (2005)
5. Jin, R., Hauptmann, A.G., Zhai, C.X.: Title language model for information retrieval. In: Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 42–48. ACM, New York (2002)
6. Jing, Y., Croft, W.B.: An association thesaurus for information retrieval. In: Proceedings of RIAO, pp. 146–160 (1994)
7. Karimzadehgan, M., Zhai, C.X.: Estimation of statistical translation models based on mutual information for ad hoc information retrieval. In: Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 323–330. ACM, New York (2010)
8. Karimzadehgan, M., Zhai, C.: Axiomatic analysis of translation language model for information retrieval. In: Baeza-Yates, R., de Vries, A.P., Zaragoza, H., Cambazoglu, B.B., Murdock, V., Lempel, R., Silvestri, F. (eds.) ECIR 2012. LNCS, vol. 7224, pp. 268–280. Springer, Heidelberg (2012)

9. Keen, E.M.: Some aspects of proximity searching in text retrieval systems. The Journal of Information Science 18(2), 89–98 (1992)
10. Lv, Y., Zhai, C.X.: Positional language models for information retrieval. In: Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 299–306. ACM, New York (2009)
11. Miao, J., Huang, X.J., Ye, Z.: Proximity-based Rocchio's model for pseudo relevance. In: Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 535–544. ACM, New York (2012)
12. Ponte, J.M., Croft, W.B.: A language modeling approach to information retrieval. In: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 275–281. ACM, New York (1998)
13. Schütze, H., Pedersen, J.O.: A co-occurrence based thesaurus and two applications to information retrieval. Information Processing and Management 33(3), 307–318 (1997)
14. Tao, T., Zhai, C.X.: An exploration of proximity measures in information retrieval. In: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 295–302. ACM, New York (2007)
15. Voorhees, E.M.: Query expansion using lexical-semantic relations. In: Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 61–69. ACM, New York (1994)
16. Zhao, J., Yun, Y.: A proximity language model for information retrieval. In: Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 291–298. ACM, New York (2009)
17. Zhao, J., Huang, X., He, B.: CRTER: Using cross terms to enhance probabilistic information retrieval. In: Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 155–164. ACM, New York (2011)
18. Zhai, C.X., Lafferty, J.: A study of smoothing methods for language models applied to Ad Hoc information retrieval. In: Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 334–342. ACM, New York (2001)
19. Zhai, C.X., Lafferty, J.: Model-based feedback in the language modeling approach to information retrieval. In: Proceedings of the Tenth International Conference on Information and Knowledge Management, pp. 403–410. ACM, New York (2001)
20. Zhai, C.X.: Statistical Language Models for Information Retrieval A Critical Review. Foundations and Trends in Information Retrieval 2(3), 137–213 (2008)