Recommending Citations: Translating Papers into References

Wenyi Huang[†]
harrywy@gmail.com

Saurabh Kataria[‡*]
saurabh.Kataria@xerox.com

Cornelia Caragea[†]
ccaragea@ist.psu.edu

Prasenjit Mitra[†]
pmitra@ist.psu.edu

C. Lee Giles[†]
giles@ist.psu.edu

Lior Rokach[§*]
liorrk@bgu.ac.il

[†]Information Sciences & Technology
The Pennsylvania State University
University Park, PA 16802

[‡]Xerox Research Center Webster
New York, US

[§]Information Systems Engineering
Ben-Gurion University of the Negev
Beer-Sheva, Israel 84105

ABSTRACT

When we write or prepare to write a research paper, we always have appropriate references in mind. However, there are most likely references we have missed and should have been read and cited. As such a good citation recommendation system would not only improve our paper but, overall, the efficiency and quality of literature search.

Usually, a citation's context contains explicit words explaining the citation. Using this, we propose a method that "translates" research papers into references. By considering the citations and their contexts from existing papers as parallel data written in two different "languages", we adopt the translation model to create a relationship between these two "vocabularies".

Experiments on both CiteSeer and CiteULike dataset show that our approach outperforms other baseline methods and increase the precision, recall and f-measure by at least 5% to 10%, respectively. In addition, our approach runs much faster in the both training and recommending stage, which proves the effectiveness and the scalability of our work.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval

General Terms

Algorithms, Experimentation

Keywords

Citation recommendation, machine translation

1 Introduction

Citations are important in academic dissemination in at least two ways. First, correct citations demonstrate intellectual honesty by giving credit to the work of others; second, proper citations help readers trace the source and evaluate whether the referenced works support authors' claims. So as to attribute completely the work of previous researchers, authors must be very careful when creating the literature review to avoid missing significant references.

*The work was done while these two authors were at Penn State University

CIKM'12, October 29–November 2, 2012, Maui, HI, USA.

Most current literature search engines focus on short queries. In our work, we mainly deal with the cases where users provide a longer query ranging from a sentence to an entire manuscript, and our recommendation system automatically suggests a list of references based on the query input.

As shown in Fig. 1, the descriptive language usually contains words that describe or summarize the main points of the cited papers. Therefore, citation recommendation can be described as a translation process, where we "translate" context sentences into papers to be cited.

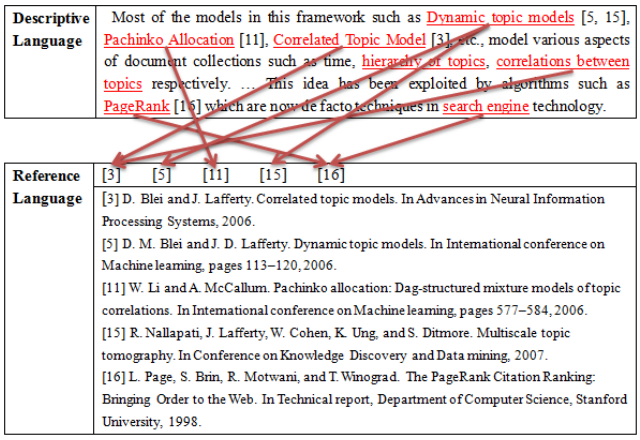


Figure 1: An example of translation from the descriptive language to the reference language, adapted from [16].

A research paper is written using two different "languages": (1) the **descriptive language**, consisting of citation words used in the paper before the reference section; and (2) the **reference language**, consisting of references, where each referenced paper is considered as a "word". In order to distinguish different papers, the reference language vocabulary is a set of unique IDs representing cited papers.

The citation translation model for reference recommendation involves two steps: (1) Build up a dictionary that contains the translation probability of a reference given a word or phrase for all terms in the descriptive language vocabulary. (2) Compute the probability of a reference given the query using the translation probabilities. Recommend references in ranked order.

The major contributions of this paper are:

- We propose to represent the cited papers by unique IDs, regarding them as "words" in a novel language, and then use translation model to estimate the translation probability of a ID given citing words. We also use the model to capture the co-citation relationship in a novel way.

- We demonstrate that our approach improves the performance by increasing the precision, recall and f-measure by at least 5% to 10%, respectively, compared with the state-of-the-art approaches.
- By comparing the model complexity with baseline methods, we show that on a large dataset our approach runs at least 100 times faster in the training stage and 5 to 600 times faster in the recommending stage, which proves the effectiveness and the scalability of our approach.

2 Related Work

2.1 Citation Recommendation

Most early works for citation recommendation require user profile information or a partial list of references. McNee, et al. [13] explore the use of collaborative filtering for recommending research papers. Their method uses the citation network, paper-citation information, and co-citation information to create a ratings matrix. This method based on the citing history of the user, however, did not take the content into consideration.

Strohman, et al. [20] introduced a citation recommendation system that uses the combination of content features and citation web information to evaluate the relevance and similarity between two documents. They construct a candidate set by first using the text similarity to select an initial set, and then adding to it all the citations of each paper in the initial set. Then, they rank all the candidate papers using features such as bibliography similarity and Katz centrality measurement.

In recent years, various citation recommendation methods have been proposed using latent topic models [2]. Nallapati, et al. [16] models the text and citation together to propose a model named link-PLSA-LDA. Link-PLSA-LDA models the cited set of paper using PLSA [8] and the citing set using the link-LDA model [6]. Kataria, et al. [9] extended the method by associating terms in the citation contexts to the cited documents. The model, cite-PLSA-LDA, assumes that the words and citations occurring in the citing paper are generated from topic-word and topic-citation multinomial distributions, respectively.

Citation context analysis has been used in information retrieval for quite some time. Previous work shows that indexing cited articles with the terms appearing in citation context can improve effectiveness of retrieval compared to indexing the whole content of cited article [19, 18]. He, et al [7] use a context-aware approach for recommending citations. This approach assumes that user has provided placeholders for citation in a query manuscript. They propose a probabilistic model to measure the relevance between documents and between the citation contexts and the document.

2.2 Translation Model

In Statistical machine translation (SMT), a document is translated according to the probability distribution $\Pr(e|f)$ that a string e in the target language is the translation of a string f in the source language.

The application of the translation model has gone far beyond simple translation. Many tasks in information retrieval and natural language processing also adopt the translation model to estimate the relationship between two different objects [1], such as sentence retrieval [15], question answering [14], and tag suggestions [11].

Lu, et al. [12] used the translation model to recommend citations. They assumed that the languages used in the citation contexts and in the cited papers' content are different, and tried to bridge these two languages by translating words in the document to words in the citation contexts. After training the model, they recommend papers according to the probability of translating a cited papers' content to a citation context. Their ranking score for recommending citations actually reveals the probability of how likely a cited paper can be summarized into a citation context.

In contrast, we propose to represent the cited papers in a concise fashion (unique IDs), regarding them as new "words" in a novel language, and we propose to directly estimate the probability of citing a paper given a citation context. Moreover, we introduce a novel way of parallelizing data that better capture the co-citation relationship such that the translation model can bridge the co-cited papers via terms appearing in the paper's citation contexts.

3 Building Up Dictionary

In this section, we will first discuss how to construct parallel training data from a given corpus, and then how to learn the translation model on the training data to build up a dictionary that captures the relationship between citations and terms in the two "languages".

3.1 Constructing Parallel Dataset

Given a corpus of research papers D_{corpus}, we divide each paper into two parts: **descriptive language** d as the source language and the corresponding **reference language** r as target language as defined in SMT, then pair these two parts as one entry within the parallel dataset.

We use the terms in the citation context to form the source language. Our preliminary experiments indicate that a fixed size window surrounding the citation mention models the cited paper better than the whole content of the citing article which is too verbose and noisy for modeling the source language. A citation context c is defined as n sentences that appear before a citation and n sentences after. Intuitively, the sentence that contains a citation is the first place where descriptive terms will appear. For example, in Fig. 1, the term "PageRank" appears right before the citation. Some of the descriptive terms can be found in nearby sentences if the writer tries to expand more details for the citation. Therefore we vary the radius of a citation context n from 1 to 3. Note that it will lose the meaning of citation context if we set the radius too large.

Suppose there are k citation contexts within a descriptive language $d = [c_1, \cdots, c_k]$ and m references within the reference language $r = [r_1, \cdots, r_m]$. We construct the parallel data by obtaining all citation contexts within a paper as source language and pairing it with all citations in the paper. Thus, one paper forms one entry for the parallel data:

$$\begin{array}{ll} \text{Source} & t_{c_1,1}, \cdots, t_{c_1,|c_1|}, \cdots, t_{c_k,1}, \cdots, t_{c_k,|c_k|} \\ & \Downarrow \\ \text{Target} & r_1, r_2, \cdots, r_m \end{array}$$

where $t_{c_i,j}$ is the jth term appearing in the ith citation context of d and r_i is the ith cited paper in r. We will refer to this method as **All-to-All** type of parallel data.

The context for neighboring citations may overlap when we set the radius to 2 or 3. We do not duplicate words in the overlap for **All-to-All** parallel data.

3.2 Learning Translation Model

After constructing the parallel data, we applied the translation model to build up a dictionary over the two vocab-

ularies. We treat both descriptive and reference language as bag of words ignoring the ordering information of both languages, so we adopt the IBM translation Model-1 [3] to learn the translation model which is most suitable for our settings.

The IBM Model-1 models the translation process based on word-level alignment. The alignment from source language $d = [t_1, \cdots, t_l]$ to target language $r = [r_1, \cdots, r_m]$ is described by a hidden variable $A = [a_1, \cdots, a_m]$. In SMT, such an alignment is interpreted as the process of translation in which two words in different languages that are aligned together share the same meaning. In the citation translation model, a word aligned to a paper indicates that the word may need that particular citation. According to an alignment A, where $a_i = j$ means r_i is aligned to t_j, the objective function for translation can be formulated as:

$$\text{Maximize} \quad \Pr(r|d) = \sum_{a_1=1}^{l} \cdots \sum_{a_m=1}^{l} \prod_{i=1}^{m} \Pr(r_i|t_{a_i})$$

$$\text{Subject to} \quad \sum_{i=1}^{m} \Pr(r_i|t_j) = 1 \qquad j = 1, 2, \cdots, l$$

where $\Pr(r_i|t_{a_i})$ is the probability of citing r_i given a term t_{a_i}, or as in SMT, the probability of translation t_{a_i} to r_i .

The objective function solved using EM algorithm [5]. Both the translation table $\Pr(r_*|t_*)$ and probabilities of all possible alignments A can be initialized with uniform distributions, the EM algorithm will iteratively calculate them until convergence. The result of the algorithm will give the model for word level recommendation probability $\Pr(r_i|t_j)$, which maximizes the translation probability of document level recommendation probability $\Pr(r|d)$.

3.2.1 Model Analysis

Null Token In the translation model, the alignment allows $a_i = 0$, indicating that an element of a target language is mapped from a *null token*. This alignment is essential for machine translation, because not all words in a target language have a specific mapping from a source language. However, in scientific papers, every citation is usually cited in the text. The citation contexts will contain terms that summarize the citation. Therefore, the alignment to a null token is meaningless in our task, so we remove such kind of mapping.

Co-citation Analysis As outlined in Section 3.1, we proposed the *All-to-All* parallel data which is a novel way to capture *co-citation* relationship. In *All-to-All* data, we pair words in all citation contexts with all references of a paper. At first glance, this pairing may seem inaccurate. However, note that citation contexts make very specific comments about the relationship of a cited paper from the perspective of the citing paper. If two papers have been co-cited within a paper, they have some connections. So the translation model built on the *All-to-All* data enables a cited paper to be modeled using terms related to co-cited papers. The more two citations co-occur, the higher the probability that the words used to describe one paper is related to the other, and, the higher the probability that they will be cited together in the future.

Take this paper for example. We cite papers from machine translation and citation recommendation. The co-occurrence of these references indicates the relationship between them. Thus, in the future when people mention the application of machine translation, they might want to cite citation recommendation papers too. Trained with *All-to-All* data, the translation model can bridge the co-cited papers via terms appearing in this paper's citation contexts.

4 Reference Recommendation Using Dictionary

After we obtain a dictionary that contains the translation table between two vocabularies in the form of triplet entries $\langle t_i, r_j, Pr(r_j|t_i)\rangle$. We can now "translate" a query into a reference list.

Given a query $Q = [t_1, \cdots, t_l]$, the task is to recommend a list of references $R = [r_1, \cdots, r_m]$. We will go through all words in Q and assign the score for each reference r_i as:

$$\Pr(r_i|Q) = \sum_{j=1}^{l} \Pr(r_i|t_j) \Pr(t_j|Q) \tag{1}$$

where $\Pr(r_i|t_j)$ is the probability of translating the term t_j to the reference r_i and $\Pr(t_j|Q)$ is the probability that the term t_j needs citations within the query.

Here we use the term-frequency-inverse-context-frequency (TF-ICF) to measure $\Pr(t_j|Q)$, the probability of a citation need. Given a query Q, TF_t is defined as the number of times a given term t appears in Q, which reveals the importance of the term t within the particular query Q. ICF gives a measure of whether the term is common or rare across all citation contexts. $\text{ICF}_t = \log \frac{|C|}{\sum_{t \in C} 1}$, where C is the set of citation contexts, and $\sum_{t \in C} 1$ indicate the number of citation contexts that contain the term t.

5 Experiments

In this section, we evaluate the performance of citation translation model on two real datasets. We use the papers' reference lists as ground truth for evaluation and compare our approach with different state-of-the-art approaches.

5.1 Datasets

The first dataset `CiteSeer` has been widely used for citation recommendation by Kataria, et al. [9], Tang and Zhang [21] and Nallapati, et al. [16]. The second dataset we use was acquired from CiteULike [1] from November 2005 to January 2008. The dataset was also used by Kataria, et al. [10] for citation recommendation. The characteristics of both datasets are shown in Table 1.

Data	D	C	W_C	R	N_c
CiteSeer	3,312	26,597	21,982	2,138	18.01
CiteULike	14,418	40,720	52,631	5,484	8.61

Table 1: D is the number of documents, C is the number of citation contexts, W_C is the number of unique words in citation contexts, R is the number of unique references, and $\bar{N}_c$ is the number of average citations a paper has.

For each dataset, we first remove the stopword and then randomly partition them into 5 subsamples and then perform a 5-fold cross validation on the exact same partition for our approach and other baseline methods.

5.2 Evaluation Metrics

Precision, Recall, F-measure For each query in the test set, we use the original set of references as the ground truth R_g. Assume that the set of recommended citations are R_r ,the correct recommedations are $R_g \cap R_r$. Precision, recall and F-measure are defined as:

$$p. = \frac{|R_g \cap R_r|}{R_r}, r. = \frac{|R_g \cap R_r|}{R_g}, f. = \frac{2p. \times r.}{p. + r.} \tag{2}$$

[1] `http://www.citeulike.org/`.

In our experiments, the number of recommended citation ranges from 1 to 20.

Precision, Recall, and F-measure evaluation do not reveal the order of recommended references. To address this problem, we select the following two additional metrics.

Binary Preference Measure (Bpref) For an query q, suppose an approach recommends a list of references S, in which the correctly recommended citations is the list R. Let r be a correct recommendation and i be an incorrect recommendation. Bpref [4] is defined as:

$$\text{Bpref} = \frac{1}{|R|}\sum_{r\in R} 1 - \frac{|i \text{ ranked higher than } r|}{|S|} \quad (3)$$

Mean Reciprocal Rank (MRR) For a query q, let rank_q be the rank of the first correct recommendation within the list. MRR [22] is defined as:

$$\text{MRR} = \frac{1}{|Q|}\sum_{q\in Q}\frac{1}{\text{rank}_q} \quad (4)$$

where Q is the testing set. MRR reveals the averaged ranking of the first correct recommendation.

5.3 Baselines and Parameter Settings

We choose to compare our approach with both context-based and not context-based approaches as follows:

- **Link-PLSA-LDA** (**link-LDA**) [16]: We turned the parameter setting as suggested in [9]. The number of topics is set to 200 for `CiteSeer` and 500 for `CiteULike`. This approach is not context based.
- **Cite-PLSA-LDA** (**cite-LDA**) [9]: We set the citation context radius n to 3 and the number of topic to 200 for `CiteSeer`, 500 for `CiteULike` which give the best results as the author suggested [9]. The approach is context-aware.
- **Context-aware Relevance Model** (**CRM**) [7]: We tuned the parameter settings as suggested in that paper. The citation context radius n is set to 3 sentences as the in Cite-PLSA-LDA model. This approach is context-aware.
- **Translation Model** (**TM**) [12]: We use GIZA++ [17] [2] to learn translation between words in citation context and words in cited paper. We tuned the parameter settings as suggested in [12]. This approach is context-aware.
- **Citation Translation Model** (**CTM**): In our method, we modify the GIZA++ toolkit [17] to learn translation probabilities using IBM Model-1. The parameters that give the best performance is the citation context radius $n = 1$, and the number of training iterations around 10.

5.4 Complexity Analysis

Denote the number of training iterations for link-LDA, cite-LDA, TM and CTM as I (I actually varies among different methods), the number of topics for link-LDA and cite-LDA as K, the average number of words each citation context has as $\bar{N_{cc}}$, the average number of words each paper has as $\bar{N_w}$, and the average citations each paper cites as $\bar{N_c}$.

For the training stage, the CRM does not need a training phase. The complexity of link-LDA is $O(IKD\cdot(\bar{N_w}+\bar{N_c}))$, cite-LDA is $O(IKD\bar{N_w})$, TM is $O(ID\bar{N_w}\bar{N_{cc}}\bar{N_c})$ and CTM is $O(ID\bar{N_{cc}}\bar{N_c}^2)$. Note that $\bar{N_c}$ is usually around 20, which is 10 to 20 times less than K (ranging from 200 to 500 or even more) and $\bar{N_{cc}}\bar{N_c} < \bar{N_w}$.

[2] GIZA++ available at: `http://code.google.com/p/giza-pp/`.

For the recommending stage, assume we have a query q with N_q terms. The complexity of link-LDA is $O(IKN_q)$, cite-LDA is $O(IKN_q)$, CRM is $O(D\bar{N_c}^2)$, TM is $O(DN_qN_w)$ and CTM is $O(N_q\bar{R_q})$, where $\bar{R_q}$ is the average number of dictionary entries for each word in q. $\bar{R_q}$ usually drops tremendously (to around 20 to 50) after several iterations if we wipe out those with too low translation probabilities.

	Training		Recommending	
	CiteSeer	CiteULike	CiteSeer	CiteULike
link-LDA	622.490s	20824.61s	1.790s	34.865s
CRM	-	-	2006.032s	3012.003s
cite-LDA	594.115s	8949.210s	1.845s	20.154s
TM	573.891s	866.227s	6287.421s	9972.11s
CTM	**53.372s**	**71.460s**	**1.480s**	**4.904s**

Table 2: Run time on `CiteSeer` and `CiteULike` dataset using parameter setting mentioned in Sec 5.3.

From Table 2 [3] and the above analysis we can see that CTM is comparatively much simpler and much more efficient for both the training and recommending tasks.

5.5 Comparing Results

For all compared methods we use the parameter settings as mentioned in Section 5.3, which give the best performance. In Figure 2, Figure 3 and Table 3, we show the results on both `CiteSeer` and `CiteULike` dataset.

	CiteSeer		CiteULike	
	Bpref	MRR	Bpref	MRR
link-LDA	0.064	0.028	0.027	0.013
CRM	0.097	0.238	0.054	0.072
cite-LDA	0.459	0.285	0.260	0.143
TM	0.422	0.288	0.393	0.285
CTM(word)	**0.645**	**0.529**	**0.627**	**0.467**

Table 3: Bpref and MRR metrics on `CiteSeer` and `CiteULike` dataset with 20 recommended paper.

From the results, we get the following observations:

First, the citation translation approach outperforms all the other baselines on both datasets across the different evaluation metrics, which showed that our approach improved the recommendation significantly and robustly. The Bpref and MRR metrics show us that the proposed method generates recommendation lists which are better ranked. The MRR results indicate that our method will recommend first correct citations with an average ranking at 2, while other baseline methods ranked first correct citations with an average ranking at 4 or even worse.

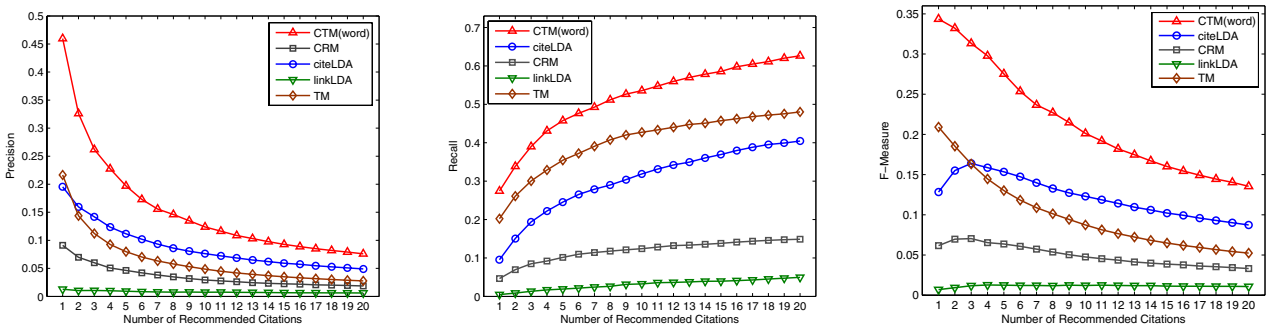
Second, as shown in Section 5.3, we have to tune the settings for cite-LDA and link-LDA according to different datasets to get a best result for each approach. For example the number of topics is set to 200 for `CiteSeer` and 500 for `CiteULike`, which was obtained empirically from experiment. Although it is intuitive that we should assign more topics for larger datasets, however, you have to train many models with different number of topics to get the best results. For the citation translation model, the only parameter needs to be tuned is the number of training iterations.

6 Conclusion and Future Work

We propose a translation-based citation recommendation model. Our approach use the existing citations and their contexts and adapted the translation model to capture mappings between terms in citation contexts and citations.

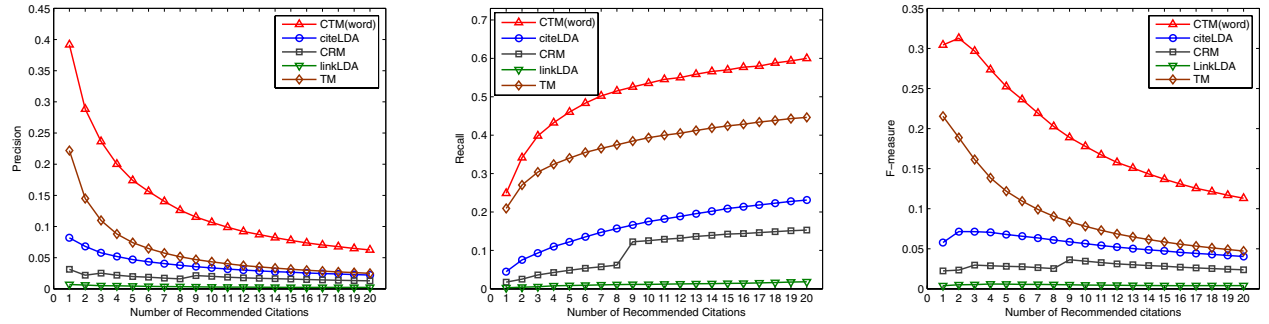
We show that using the citation contexts of all citations in a document together as the source language and the set

[3] Experiments were conducted on a same machine with 8 cpus processors of 2.50GHz and 32G memory.



(a) Precision (b) Recall (c) F-measure

Figure 2: Precision, recall and F-measure of different methods on `CiteSeer` dataset with recommended citations range from 1 to 20.



(a) Precision (b) Recall (c) F-measure

Figure 3: Precision, recall and F-measure of different methods on `CiteULike` dataset with recommended citations range from 1 to 20.

of references in the document as the target language captures co-citation and improves the quality of recommendation. Experiments on two real datasets demonstrated that the proposed translation approach outperforms the existing state-of-the-art methods.

We plan to investigate the following problems:

- CTM can only recommend citations that have been cited before. For newly published papers, it is hard to recommend them if they have not been cited. We plan to incorporate summarization and keyword extraction techniques to help put non-cited papers into translation tables.
- Different authors may cite different papers according to personal preferences or different emphases. Our approach is author-oblivious. We might obtain improved performance when the authors are taken into consideration.

7 References

[1] A. Berger and J. Lafferty. Information retrieval as statistical translation. In *Proc. of SIGIR'99*, pages 222–229. ACM, 1999.

[2] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, pages 993–1022, 2003.

[3] P. F. Brown, V. J. D. Pietra, S. A. D. Pietra, and R. L. Mercer. The mathematics of statistical machine translation: parameter estimation. *Comput. Linguist.*, 19:263–311.

[4] C. Buckley and E. Voorhees. Retrieval evaluation with incomplete information. In *Proc. of SIGIR'04*, pages 25–32, 2004.

[5] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B*, pages 1–38, 1977.

[6] E. Erosheva, S. Fienberg, and J. Lafferty. Mixed membership models of scientific publications. In *Proc. of the National Academy of Sciences*, 2004.

[7] Q. He, J. Pei, D. Kifer, P. Mitra, and C. L. Giles. Context-aware citation recommendation. In *Proc. of WWW'10*, pages 421–430. ACM, 2010.

[8] T. Hofmann. Probabilistic latent semantic indexing. In *Proc. of SIGIR'99*, pages 50–57. ACM, 1999.

[9] S. Kataria, P. Mitra, and S. Bhatia. Utilizing context in generative bayesian models for linked corpus. In *Proc. of AAAI'10*, 2010.

[10] S. Kataria, P. Mitra, C. Caragea, and C. L. Giles. Context sensitive topic models for author influence in document networks. In *Proc. of IJCAI'11*, pages 2274–2280, 2011.

[11] Z. Liu, X. Chen, and M. Sun. A simple word trigger method for social tag suggestion. In *Proc. of EMNLP'11*. ACL, 2011.

[12] Y. Lu, J. He, D. Shan, and H. Yan. Recommending citations with translation model. In *Proc. of CIKM'11*, pages 2017–2020. ACM, 2011.

[13] S. M. McNee, I. Albert, D. Cosley, P. Gopalkrishnan, S. K. Lam, A. M. Rashid, J. A. Konstan, and J. Riedl. On the recommending of citations for research papers. In *Proc. of CSCW'02*, pages 116–125. ACM, 2002.

[14] V. Murdock. Simple translation models for sentence retrieval in factoid question answering. In *Proc. of SIGIR'04*, pages 31–35, 2004.

[15] V. Murdock and W. B. Croft. A translation model for sentence retrieval. In *Proc. of HLT/EMNLP*, HLT '05, pages 684–691. ACL, 2005.

[16] R. M. Nallapati, A. Ahmed, E. P. Xing, and W. W. Cohen. Joint latent topic models for text and citations. In *Proc. of SIGKDD'08*, pages 542–550. ACM, 2008.

[17] F. J. Och and H. Ney. Improved statistical alignment models. In *Proc. of ACL'00*, 2000.

[18] A. Ritchie, S. Robertson, and S. Teufel. Comparing citation contexts for information retrieval. In *Proc. of CIKM'08*, pages 213–222. ACM, 2008.

[19] A. Ritchie, S. Teufel, and S. Robertson. Using terms from citations for ir: some first results. In *Proc. of ECIR'08*, pages 211–221. Springer-Verlag, 2008.

[20] T. Strohman, W. B. Croft, and D. Jensen. Recommending citations for academic papers. In *Proc. of SIGIR'07*, pages 705–706. ACM, 2007.

[21] J. Tang and J. Zhang. A discriminative approach to topic-based citation recommendation. In *Proc. of PAKDD'09*, pages 572–579. Springer-Verlag, 2009.

[22] E. Voorhees. The trec-8 question answering track report. In *Proc. of TREC'00*, pages 77–82, 2000.