A Submodular Approach for Reference Recommendation

Thanh-Binh Kieu
Joint Technology and Innovation Research Centre
University of Technology Sydney - Vietnam National University
Hanoi, Vietnam
Email: binhkt@vnu.edu.vn

Xuan-Hieu Phan
Faculty of Information Technology
University of Engineering and Technology, VNU
Hanoi, Vietnam
Email: hieupx@vnu.edu.vn

Bao-Son Pham
Vietnam National University
Hanoi, Vietnam
Email: sonpb@vnu.edu.vn

Massimo Piccardi
School of Electrical and Data Engineering
Faculty of Engineering and IT, UTS
Broadway NSW 2007, Australia
Email: Massimo.Piccardi@uts.edu.au

Abstract—Choosing appropriate references for a given topic is an important, yet challenging task. The pool of potential candidates is typically very large, in the order of tens of thousands, and growing by the day. For this reason, this paper proposes an approach for automatically providing a reference list for a given abstract. The approach is based on an original submodular inference function which balances relevance, coverage and diversity in the reference list. Experiments are carried out using an ACL corpus as a source for the references and evaluated in terms of precision-recall, MAP and MRR. The results show the remarkable comparative performance of the proposed approach.

Keywords-Reference recommendation, submodular inference, monotonic submodular functions.

I. INTRODUCTION

Have you ever been overwhelmed by the large number of research papers in your research field? Are you wondering which ones you should read for your research, and afraid of missing some important, new ideas? Does it feel like the deluge is only getting worse?

To ease this challenge, this paper proposes a novel model, named **SubRef**, that can recommend an effective and adequate reference list for a user-provided query. The reference list is predicted based on a combination of relevance, coverage and diversity. Specifically, the relevance refers to the relevance of the recommended articles to the user query, which can be computed in terms of similarity between the query and the articles' contents; the coverage reflects the extent to which the recommended list is able to cover the query; and the enforced diversity across the recommended articles prevents redundancies and overlapping within the list. In addition, by its nature, the proposed model lends itself to accommodate specialized recommendations (for instance, for a given user profile: e.g., newcomer vs. expert).

The ability to provide reference recommendations that meet out three stated criteria stems from the powerful domain of submodular inference [1]. Submodular functions are capable of encapsulating many desirable properties of subsets while allowing for fast inference of good subsets with theoretical performance guarantees. As such, they are natural candidates for the selection of effective reference recommendations from large corpora of documents.

The main contributions of our paper are summarized as follows.

- We propose a novel model that recommends relevant, covering and diverse references which can help, amongst other, in the writing of scientific papers.
- We propose a class of effective submodular functions to be used for the provision of reference recommendations. We also analyze their time complexity and provide a comparative performance evaluation.
- Thorough experiments are carried out using the ACL Anthology Reference Corpus to validate the effectiveness of the proposed model.

II. RELATED WORK

Reference recommendation approaches can be divided into two main categories based on different styles of querying: global and inline. The global recommendation approaches predict references that are relevant to an entire manuscript. They typically employ the whole manuscript or significant parts thereof such as the title, abstract, author and venue as the query [2]. By contrast, inline, or contextual, reference recommendation approaches use the local context of each "placeholder" to capture its specific information requirements [3]. Although our approach can apply to both styles of reference recommendation, we only focus on global reference recommendation in the following.

From the perspective of the underlying technology, recommendation approaches can instead be divided into three main groups: collaborative filtering (CF) approaches, graph-based (GB) approaches and content-based filtering (CBF)

approaches. Each group has its own rationale for basing the recommendations: CF mainly focuses on the recommendations or ratings of other users whose profiles are similar to the querying user [4]. CBF calculates the similarity between keywords extracted from the user's query and from candidate papers [5], [6]. In turn, GB methods construct a graph in which authors and papers (possibly including venues and other meta-information) are regarded as nodes, and the edges express paper-to-paper [7], author-to-author, and paper-to-author relationships. Recommendations are eventually provided in terms of optimal graph traversals. In addition, various combinations of these three groups, generically referred to as *hybrid methods*, have also been proposed to improve the accuracy of the recommendation results and obtain the better performance [8], [9].

A submodular approach to reference recommendation has also been proposed in the literature [10]. To mollify computational issues, it introduces a streaming algorithm with a constant-factor approximation guarantee that uses only a limited amount of memory. Differently from [10], in this paper we propose a novel class of submodular functions of limited computational cost which is capable of providing effective recommendation lists out of large candidate pools.

III. Submodularity Background

A. Definitions

To clarify the connection between submodularity and reference recommendation, hereafter we recap the main properties of submodularity.

Definition 1 (Submodularity) A function $f : 2^V - > R$ is *submodular* if for every $A, B \subseteq V$ it holds that
$$f(A \cap B) + f(A \cup B) \leq f(A) + f(B)$$
Equivalently, a function $f : 2^V - > R$ is *submodular* if for every $A \subseteq B \subseteq V$ and $e \in V$ it holds that
$$\Delta(e \mid A) := (f(A \cup e) - f(A)) \geq \Delta(e \mid B).$$
Intuitively, this definition says that the benefit achieved for adding element e to an existing summary is greater or equal if such a summary is smaller.

An important subclass of submodular functions are those which are *monotone*, where enlarging the argument set cannot cause the function to decrease:

Definition 2 (Monotonicity) A function $f : 2^V - > R$ is *monotone* if for every $A \subseteq B \subseteq V, f(A) \leq f(B)$.

The concept of submodularity fits reference selection in recommendation tasks well: in this case, V is the set of all the candidate references, e is an element in V, and A and B are two recommendation lists (i.e., subsets of V). Intuitively, there will be less "gain" for introducing another reference into a list if such a list is already substantial. Therefore, our aim becomes that of maximizing submodular functions, i.e. finding $\max_{S \subseteq V} f(S)$, subject to some constraints on S. The simplest type of constraint is the cardinality constraint, where we require that $|S| \leq k$, with k an acceptable number

of references for a scientific publication, typically in the order of 20 to 30.

The problem of maximizing submodular functions is NP-hard and usually approximately solved via a simple, greedy algorithm which, however, enjoys theoretical guarantees for its worst-case approximation. The greedy algorithm is provided below.

The Greedy Algorithm (Algorithm 1) With S_0 the empty set, at iteration $i = 1 \dots k$, add element $d \in V \setminus S_{i-1}$ maximizing the discrete derivative $\Delta(d \mid S_{i-1})$:

$$S_i = S_{i-1} \cup \{\text{argmax}_d \Delta(d \mid S_{i-1})\} \tag{1}$$

Algorithm 1 Greedy submodular function maximization

1: $S^* \leftarrow \emptyset$
2: $A \leftarrow V = \{d_1, d_2, ...\}$
3: **while** $A \neq \emptyset$ **and** $|S^*| < k$ **do**
4: $\quad z \leftarrow \text{argmax}_{d \in A} F(S^* \cup d) - F(S^*)$
5: $\quad S^* \leftarrow S^* \cup \{z\}$
6: $\quad A \leftarrow A \setminus \{z\}$
7: **end while**

A celebrated result by Nemhauser et al. (1978) [1] proves that the greedy algorithm provides a good approximation to the optimal solution of the NP-hard optimization problem.

Theorem 1 (Nemhauser et al. 1978) Given a nonnegative monotone submodular function $f : 2^V - > R_+$, let $S_1, S_2 \dots S_k$ be the greedily selected sets defined in Eq. 1. Then, for any positive integer $l \leq k$,

$$f(S_l) \geq (1 - e^{-l/k}) \max_{S:|S| \leq k} f(S) \tag{2}$$

In particular, for $l = k$, $f(S_k) \geq (1 - 1/e) \max_{S:|S| \leq k} f(S)$.

B. Submodular functions used in document summarization

Document summarization is a well-studied problem in the literature, with interesting analogies with reference recommendation. Document summarization is often solved at sentence level, whereas reference recommendation is solved at document level with each document as a reference. Hereafter, we describe the main submodular functions for document summarization that can form the basis for our proposal.

Lin and Bilmes [11] have been the first to frame document summarization as the maximization of a submodular function under a budget constraint. They used an objective consisting of a coverage term combined with a penalty for redundancy:

$$f_{MMR}(S) = \sum_{i \in V-S} \sum_{j \in S} w_{ij} - \lambda \sum_{i,j \in S:i \neq j} w_{ij}, \lambda \geq 0. \tag{3}$$

in which $S \subset V$ and w_{ij} is the similarity score between sentence i and sentence j. Please note that this scoring

function is not guaranteed monotone: while the coverage term can be easily proved to be monotone, the penalty term may surpass it and make increments of (3) become negative. However, a monotone behavior can still be expected for reasonably small summaries.

In [12], the same authors studied a class of monotone submodular functions that combine two terms: one which encourages the summary to be representative of the corpus and the other which positively rewards diversity. The summary score is modeled as:

$$F(S) = F_c(S) + \lambda F_d(S) \tag{4}$$

where $F_c(S)$ measures coverage and $F_d(S)$ rewards diversity. In turn:

$$F_c(S) = \sum_{i \in V} min\{C_i(S), \alpha C_i(S)\} \tag{5}$$

where $C_i(S)$ measures of well element i is covered by summary S, which "saturates" when it reaches a given fraction, α, of its largest possible value, $C_i(V)$; and:

$$F_d(S) = \sum_{i=1}^{K} \sqrt{\sum_{j \in P_i \cap S} r_j} \tag{6}$$

where $P_i, i = 1, \ldots K$ is a partition of the original set, V, and $r_j \geq 0$ indicates the reward of adding i to the empty set. Since the square root grows less than linearly, this penalty favors selecting summary elements from different clusters.

Other studies have used a very similar submodular framework, but slightly different design and analysis. Among them, Dasgupta et al. [13] have formulated the objective function as a sum of a submodular function and a non-submodular function called dispersion, with the latter using inter-sentence dissimilarities in different ways while targeting non-redundancy of the summary.

IV. SUBMODULAR REFERENCE RECOMMENDATION

Two very desirable properties of a good reference list are *relevance* and *non-redundancy*. Traditional recommendation systems usually measure relevance by ranking candidate recommendations from the most relevant to the least relevant based on a query. On the other hand, the submodular summarization approaches reviewed in the previous section provide summaries that are query-independent. For this reason, we propose extending the submodular scoring functions with an additional query-dependent *relevance* term, and we prove that the resulting function is capable of retaining submodularity.

A. Non-monotone Submodular Functions

The relevance term that we propose is defined as:

$$F_q(S) = \sum_{i \in S} sq_i \tag{7}$$

where sq_i is the similarity score between the given query, q, and document i in the summary.

Theorem 2 (original) We prove that (7) is submodular by this simple argument: given a summary, S, the score increment provided by adding a new document, d, to it only depends on the document itself and not on the summary. Therefore, (7) satisfies the second non-strict inequality in Definition 1 with the equal sign. In addition, any convex combination of (7) with other submodular functions is submodular by construction.

Our first non-monotone submodular function considers two terms: relevance and non-redundancy. When adding a new element, k, to an existing summary, S, the increment of this function can be expressed as:

$$\Delta(k) = \lambda sq_k - (1 - \lambda) \max_{i \in S} w_{ik} \tag{8}$$

The second non-monotone submodular function considers relevance, coverage and non-redundancy:

$$F(S, q) = \alpha F_q(S) + F_{cr}(S) \tag{9}$$

where $F_q(S)$ is as in (7), and:

$$F_{cr}(S) = \lambda \sum_{i \in V \setminus S} \sum_{j \in S} w_{ij} - (1 - \lambda) \sum_{i,j \in S} w_{ij} \tag{10}$$

As we said above, w_{ij} is the similarity score between document i and document j. In the simplest case, w_{ij} can be set to be the TFIDF cosine similarity [14], In alternative, the BM25 score [15] is able provide more sophisticated similarity measurements and rankings.

B. Monotone Submodular Functions

To also encapsulate information about the authors and references, we propose a class of functions which combine the relevance to the query with the relations between authors and references:

$$F(S) = \sum_{i=1}^{K} \sqrt{\sum_{j \in P_i \cap S} R(d_j)} \tag{11}$$

$$R(d_i) = \lambda F_q(d_i) + (1 - \lambda) F_{au}(d_i) \tag{12}$$

where:

$$F_{au}(d_i) = \sum_{Au(d_i) \cap Au(q) \neq \emptyset} \omega_1 + \sum_{\substack{d_j \in cite(d_i), \\ Au(d_j) \cap Au(q) \neq \emptyset}} \omega_2 \tag{13}$$

Year	2012
Papers	1186
Authors	1657
Venues	34
References	11631
Avg Cite	9.8

Table I
MAIN STATISTICS FOR THE TEST SET.

with $cite(d_i)$ being the reference list of document i, $Au(d_i)$ its set of authors, and ω_1, ω_2 the scores credited for sharing authors and references, respectively. With a preliminary evaluation, we have chosen $\omega_1 = 2.0$ and $\omega_2 = 1.0$. This function is submodular and also naturally monotone since it does not include any penalty.

C. Complexity Analysis

V. EXPERIMENTS

A. Corpus

In order to validate our model for reference recommendation, we use the ACL Anthology Reference Corpus (AAN) which was established by [16], a set of 22,085 papers from computational linguistics workshops, conferences and journals. We remove the papers which do not have titles or abstracts in the data set, then we use the papers from 1965 to 2012 as the experimental data set. For evaluation purposes, we use all papers in 2012 containing ACL references for the test set. Table I summarizes the statistics of the dataset.

For our retrieval experiments, we first construct a query by an article's title and abstract; then, we let our model provide the recommended references for this query and compare the results with the actual reference list of the article.

B. Evaluation Metrics

To evaluate the quality of the reference recommendations, we use the reference lists of the test papers as the ground truth. Following common practice, we employ the following evaluation metrics: Precision and Recall, which are the most commonly used metrics in the reference recommendation field (see survey [8]); Precision@N (P@N) which measures the percentage of the retrieved references that is relevant to the ground truth in the top-N recommendation list; Recall@N (R@N) which measures the proportion of the actual references that are retrieved in the top-N recommendation list; and other derived metrics. These metrics are calculated as follows:

$$Precision = \frac{\sum_{d \in Q} |R(d) \cap T(d)|}{\sum_{d \in Q} |R(d)|} \quad (14)$$

$$Recall = \frac{\sum_{d \in Q} |R(d) \cap T(d)|}{\sum_{d \in Q} |T(d)|} \quad (15)$$

where Q is the test set of papers, $T(d)$ are the ground-truth references of paper d, and $R(d)$ are its recommended references.

In addition, the Average Precision (AP) computes a precision-recall trade-off by calculating the precision at each point in the recommended list where a ground-truth article appears [17]. The AP can be expressed as:

$$AP(d) = \sum_{i \in [1, |R(d)|]} \frac{Precision(d)@N(i)}{|T(d)|} \quad (16)$$

where $Precision(d)$ is the precision for document d and $N(i)$ is the minimum number of top recommendations required to include reference i. In turn, the Mean Average Precision (MAP) averages the AP value over the entire test set:

$$MAP = \sum_{d \in Q} \frac{AP(d)}{|Q|} \quad (17)$$

Eventually, the Mean Reciprocal Rank (MRR) is the average of the reciprocal ranks of the recommendations for all documents in the test set. The reciprocal rank of a recommendation is the inverse of the rank position of the first correct recommendation for a test paper. The MRR is calculated as:

$$MRR = \frac{1}{|Q|} \sum_{d \in Q} \frac{1}{rank_d} \quad (18)$$

C. Experimental Settings

For a comprehensive comparison, we have evaluated the proposed model in a number of variants together with various baselines and a state-of-the-art method:

ES-TFIDF: The similarity score based on TFIDF is a common method to search related documents. For every query in the test set, we use it to rank all the documents and select the top N.

ES-BM25: BM25 is a well-known ranking method for measuring the relevance of documents to a query based on their content. The newest version of the popular Elasticsearch (ES)[1] uses BM25 as its default similarity score. The authors state that there can be significant advantages in using BM25 over TFIDF as similarity measurement, at least in some cases [18]. For this reason, we use BM25 for calculating the similarity score between the query and the documents. In the experiments, we set bias term b to 0.75 and bias term k_1 to 1.2 (the ES default).

TopicCite: TopicCite is a state-of-the-art reference recommendation approach which leverages a joint model of feature regression and topic learning. [2].

SubRef-QFRv1: our model as per Equation 8. As values for λ we have used range $(0.0, 1.0)$ in 0.1 steps. $\lambda = 0.0$

---

[1] https://www.elastic.co/

Lambda	No.Correct	MRR	MAP	P@100	R@100
0.0	3674	0.4418	0.1228	0.0310	0.3159
0.1	3783	0.4433	0.1270	0.0319	0.3253
0.2	3947	0.4461	0.1318	0.0333	0.3394
0.3	4101	0.4495	0.1358	0.0346	0.3526
0.4	4265	0.4529	0.1412	0.0360	0.3666
0.5	4469	0.4569	0.1472	0.0377	0.3842
0.6	4629	0.4608	0.1523	0.0390	0.3980
0.7	4765	0.4649	0.1573	0.0402	0.4097
0.8	**4820**	0.4715	**0.1594**	**0.0406**	**0.4144**
0.9	4766	**0.4739**	0.1580	0.0402	0.4098
1.0	4585	0.4729	0.1516	0.0386	0.3942

Table II
SUBREF-QFRV1 BY LAMBDA

Lambda	No.Correct	MRR	MAP	P@100	R@100
0.0	826	0.3697	0.0601	0.0070	0.0710
0.1	1170	0.3778	0.0629	0.0099	0.1006
0.2	1806	0.3922	0.0703	0.0152	0.1553
0.3	2490	0.4093	0.0811	0.0210	0.2141
0.4	3064	0.4264	0.0947	0.0258	0.2634
0.5	3526	0.4406	0.1084	0.0297	0.3031
0.6	3917	0.4508	0.1219	0.0330	0.3368
0.7	4194	0.4592	0.1329	0.0354	0.3606
0.8	4356	0.4672	0.1413	0.0367	0.3745
0.9	4489	0.4700	0.1475	0.0378	0.3859
1.0	**4585**	**0.4729**	**0.1516**	**0.0386**	**0.3942**

Table III
SUBREF-QFRV2 BY LAMBDA

Lambda	No.Correct	MRR	MAP	P@100	R@100
0.0	4265	0.4612	0.1281	0.0360	0.3667
0.1	4844	0.4885	0.1431	0.0408	0.4165
0.2	4913	**0.5073**	0.1521	0.0414	0.4224
0.3	4936	0.5036	0.1559	0.0416	0.4244
0.4	4938	0.4944	0.1550	0.0416	0.4246
0.5	4937	0.4772	0.1543	0.0416	0.4244
0.6	4936	0.4576	0.1497	0.0416	0.4244
0.7	4936	0.4418	0.1462	0.0416	0.4243
0.8	4936	0.4297	0.1445	0.0416	0.4244
0.9	4937	0.4277	0.1433	0.0378	0.4245
1.0	**5192**	0.4375	**0.1565**	**0.0438**	**0.4464**

Table IV
SUBREF-QAIV1 BY LAMBDA

Lambda	No.Correct	MRR	MAP	P@100	R@100
0.0	4585	0.4729	0.1516	0.0386	0.3942
0.1	4626	0.4754	0.1529	0.0390	0.3977
0.2	4671	0.4778	0.1543	0.0393	0.4016
0.3	4728	0.4809	0.1563	0.0399	0.4065
0.4	4797	0.4833	0.1591	0.0404	0.4124
0.5	4868	0.4919	0.1629	0.0410	0.4185
0.6	4946	0.5017	0.1687	0.0417	0.4252
0.7	5036	0.5164	0.1742	0.0425	0.4330
0.8	5131	**0.5179**	**0.1798**	0.0433	0.4411
0.9	**5192**	0.4949	0.1787	**0.0438**	**0.4464**
1.0	5190	0.4374	0.1564	0.0437	0.4463

Table V
SUBREF-QAIV2 BY LAMBDA

search for false negatives.

D. Parameter Tuning

In this section, we study the impact of parameters α and λ in our non-monotone submodular functions. In the QFRv1 function, λ is the trade-off between relevance and non-redundancy. In the QFRv2 function, α is the trade-off between relevance and a coverage and non-redundancy combination, while λ trades off coverage and non-redundancy. In functions QAIv1 and QAIv2, λ is the trade-off between the relevance and the importance of the author clusters. As values, for α we have empirically chosen 5.0 based on preliminary trials. For λ, we report all results in range [0.0,1.0] in 0.1 steps. In turn, the trade-off between coverage and diversity in Equation 11 is encapsulated by the square root term.

E. Performance Comparison

All results are shown in Tables II to VI. It is evident that the submodular functions with suitably-chosen parameters significantly outperform the baseline methods *in all cases*. The state-of-the-art method TopicCite still outperforms our method in the available metrics. However, TopicCite is a fully trained method of signficant computational cost, while our method *does not require any training* and runs in linear time. All experiments have been carried out on a high-performance computer with a 12-core Intel Xeon(R) CPU E5-2697 v4 @ 2.30GHz with an 8 x 16 GB DIMM ECC DDR4 @ 2400MHz RAM.

VI. CONCLUSION

In this paper, we have presented a novel approach for submodular inference of reference recommendations, an important information retrieval application. Our experiments have proven that submodular functions that had first been proposed for document summarization also give significant benefits in reference recommendation and are able to outperform relevant baselines. Although our results have not yet overcome state-of-the-art methods, there are many potential directions for further developing this work, including

means that the scoring function only scores non-redundancy and $\lambda = 1.0$ means that function only scores query relevance, which makes it equivalent to the baseline methods.

SubRef-QFRv2: our model as per Equation 9. We have used $\alpha = 5.0$ and λ as above.

SubRef-QAIv1: our model as per Equation 11, with partition P obtained by clustering the venues.

SubRef-QAIv2: our model as per Equation 11, with partition P obtained by clustering the authors.

For all comparisons, we use budget $k = 100$. This value is rather large and is chosen to favor recall over precision, in the assumption that the average user would prefer a brief manual refinement of an abundant list over an unbounded

Method	Budget	MRR	MAP	P@100	R@100
ES-TFIDF	100	0.2768	0.0640	0.0178	0.1814
ES-BM25	100	0.4729	0.1516	0.0386	0.3942
TopicCite	Unknown	**0.5713**	Unknown	Unknown	**0.5035**
SubRef-QFRv1	100	0.4739	0.1594	0.0406	0.4144
SubRef-QFRv2	100	0.4700	0.1475	0.0378	0.3860
SubRef-QAIv1	100	0.5073	0.1565	0.0438	0.4464
SubRef-QAIv2	100	**0.5179**	**0.1798**	0.0438	0.4464

Table VI
PERFORMANCE COMPARISON BETWEEN DIFFERENT METHODS.

integrating a step of submodular inference into existing methods and training the parameters automatically against loss functions. We plan to incorporate these features in our approach in the near future.

REFERENCES

[1] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximizing submodular set functions–i," *Math. Program.*, vol. 14, no. 1, pp. 265–294, Dec. 1978. [Online]. Available: https://doi.org/10.1007/BF01588971

[2] T. Dai, L. Zhu, Y. Wang, H. Zhang, X. Cai, and Y. Zheng, "Joint model feature regression and topic learning for global citation recommendation," *IEEE Access*, vol. 7, pp. 1706–1720, 2019. [Online]. Available: https://doi.org/10.1109/ACCESS.2018.2884981

[3] Q. He, J. Pei, D. Kifer, P. Mitra, and L. Giles, "Context-aware citation recommendation," in *Proceedings of the 19th International Conference on World Wide Web*, ser. WWW '10. New York, NY, USA: ACM, 2010, pp. 421–430. [Online]. Available: http://doi.acm.org/10.1145/1772690.1772734

[4] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry, "Using collaborative filtering to weave an information tapestry," *Commun. ACM*, vol. 35, no. 12, pp. 61–70, Dec. 1992. [Online]. Available: http://doi.acm.org/10.1145/138859.138867

[5] F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, *Recommender Systems Handbook*, 1st ed. Berlin, Heidelberg: Springer-Verlag, 2010.

[6] Y. Ding, G. Zhang, T. Chambers, M. Song, X. Wang, and C. Zhai, "Content-based citation analysis: The next generation of citation analysis," *Journal of the Association for Information Science and Technology*, vol. 65, no. 9, pp. 1820–1833, 2014. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.23256

[7] Y. Liang, Q. Li, and T. Qian, "Finding relevant papers based on citation relations," in *Proceedings of the 12th International Conference on Web-age Information Management*, ser. WAIM'11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 403–414. [Online]. Available: http://dl.acm.org/citation.cfm?id=2035562.2035609

[8] X. Bai, M. Wang, I. Lee, Z. Yang, X. Kong, and F. Xia, "Scientific paper recommendation: A survey," *IEEE Access*, vol. 7, pp. 9324–9339, 2019.

[9] J. Beel, B. Gipp, S. Langer, and C. Breitinger, "Research-paper recommender systems: A literature survey," *Int. J. Digit. Libr.*, vol. 17, no. 4, pp. 305–338, Nov. 2016. [Online]. Available: http://dx.doi.org/10.1007/s00799-015-0156-0

[10] Q. Yu, E. L. Xu, and S. Cui, "Submodular maximization with multi-knapsack constraints and its applications in scientific literature recommendations," in *2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Dec 2016, pp. 1295–1299.

[11] H. Lin and J. Bilmes, "Multi-document summarization via budgeted maximization of submodular functions," in *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Los Angeles, California: Association for Computational Linguistics, Jun. 2010, pp. 912–920. [Online]. Available: https://www.aclweb.org/anthology/N10-1134

[12] ——, "A class of submodular functions for document summarization," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*. Portland, Oregon, USA: Association for Computational Linguistics, Jun. 2011, pp. 510–520. [Online]. Available: https://www.aclweb.org/anthology/P11-1052

[13] A. Dasgupta, R. Kumar, and S. Ravi, "Summarization through submodularity and dispersion," in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Sofia, Bulgaria: Association for Computational Linguistics, Aug. 2013, pp. 1014–1022. [Online]. Available: https://www.aclweb.org/anthology/P13-1100

[14] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Inf. Process. Manage.*, vol. 24, no. 5, pp. 513–523, Aug. 1988. [Online]. Available: http://dx.doi.org/10.1016/0306-4573(88)90021-0

[15] S. E. Robertson and S. Walker, "Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval," in *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '94. New York, NY, USA: Springer-Verlag New York, Inc., 1994, pp. 232–241. [Online]. Available: http://dl.acm.org/citation.cfm?id=188490.188561

[16] D. R. Radev, P. Muthukrishnan, V. Qazvinian, and A. Abu-Jbara, "The acl anthology network corpus," *Language Resources and Evaluation*, pp. 1–26, 2013. [Online]. Available: http://dx.doi.org/10.1007/s10579-012-9211-2

[17] R. A. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1999.

[18] S. Robertson and H. Zaragoza, "The probabilistic relevance framework: Bm25 and beyond," *Found. Trends Inf. Retr.*, vol. 3, no. 4, pp. 333–389, Apr. 2009. [Online]. Available: http://dx.doi.org/10.1561/1500000019