An Axiomatic Framework for Result Diversification*

Sreenivas Gollapudi
Microsoft Search Labs, Microsoft Research
sreenig@microsoft.com

Aneesh Sharma
Institute for Computational and Mathematical Engineering, Stanford University
aneeshs@stanford.edu

Abstract

Informally speaking, diversification of search results refers to a trade-off between relevance and diversity in the set of results. In this article, we present a unifying framework for search result diversification using the axiomatic approach. The characterization provided by the axiomatic framework can help design and compare diversification systems, and we illustrate this using several examples. We also show that several diversification objectives can be reduced to the combinatorial optimization problem of facility dispersion. This reduction results in algorithms with provable guarantees for a number of well-known diversification objectives.

1 Introduction

In present day search engines, a user expresses her information need with as few query terms as possible. In such a scenario, a small number of terms often specify the intent in only an implicit manner. In the absence of explicit information representing user intent, the search engine needs to "guess" the results that are most likely to satisfy different intents. In particular for an ambiguous query such as *eclipse*, the search engine could either take the probability ranking principle approach of taking the "best guess" intent and showing the results, or it could choose to present search results that maximize the probability that a user with a random intent finds *at least one* relevant document on the results page. This problem of the user not finding any *any* relevant document in her scanned set of documents is defined as *query abandonment*. Result diversification lends itself as an effective solution to minimizing query abandonment [1, 7, 16].

Intuitively, result diversification implies a trade-off between having more relevant results of the "most probable" intent and having diverse results in the top positions for a given query[4, 6]. The twin objectives of being diverse and being relevant often compete with each other, and any result diversification system must figure out how to trade-off these two objectives appropriately. Therefore, result diversification can be viewed as combining both ranking (presenting more relevant results in the higher positions) and clustering (grouping document satisfying similar intents) and therefore addresses a loosely defined goal of picking a set of most relevant but

Bulletin of the IEEE Computer Society Technical Committee on Data Engineering

*An earlier version appeared in the Proceedings of the 18th International Conference on World Wide Web, 2009 [8].

novel documents. This resulted in the development of a set of very different objective functions and algorithms ranging from combinatorial optimizations [4, 16, 1] to those based on probabilistic language models [6, 20]. The underlying principles supporting these techniques are often different and therefore admit different trade-off criteria. Given the importance of the problem there has been relatively little work aimed at understanding result diversification independent of the objective functions or the algorithms used to solve the problem.

This article discusses an axiomatic framework for result diversification. We begin with a set of simple and natural properties that any diversification system ought to satisfy and these properties help serve as a *basis* for the space of objective functions for result diversification. We then analyze a few objective functions that satisfy different subsets of these properties. Generally, a diversification function can be thought of as taking two application specific inputs *viz,* a relevance function that specifies the relevance of document for a given query, and a distance function that captures the pairwise similarity between any pair of documents in the set of relevant results for a given query. In the context of web search, one can use the search engine's ranking function[1] as the relevance function. The characterization of the distance function is not that clear. In fact, designing the right distance function becomes a central factor to an effective result diversification. For example, by restricting the distance function to a *metric* by imposing the triangle inequality $d(u, w) \leq d(u, v) + d(v, w)$ for all $u, v, w \in U$, we can exploit efficient approximation algorithms to solve certain class of diversification objectives (see Section 3).

Our work is similar in spirit to earlier work on axiomatization of ranking and clustering systems [3, 11]. We study the functions that arise out of the requirement of satisfying a set of simple properties and show an *impossibility result* which states that there exists no diversification function f that satisfies all the properties. We state the properties in Section 2.

Although we do not aim to completely map the space of objective functions in this study, we show that some diversification objectives reduce to different versions of the well-studied *facility dispersion* problem. Specifically, we pick three functions that satisfy different subsets of the properties and characterize the solutions obtained by well-known approximation algorithms for each of these functions. We also characterize some of the objective functions defined in earlier works [1, 16, 6] using the axioms.

1.1 Related Work

The early work of [4] described the trade-off between relevance and novelty via the choice of a parameterized objective function. Subsequent work on *query abandonment* by [6] is based on the idea that documents should be selected sequentially according to the probability of the document being relevant *conditioned* on the documents that come before. [16] solved a similar problem by using bypass rates of a document to measure the overall likelihood of a user bypassing all documents in a given set. Thus, the objective in their setting was to produce a set that minimized likelihood of completely getting bypassed.

[1] propose a diversification objective that tries to maximize the likelihood of finding a relevant document in the top-k positions given the categorical information of the queries and documents. Other works on topical diversification include [18, 21]. [20, 19] propose a risk minimization framework for information retrieval that allows a user to define an arbitrary *loss* function over the set of returned documents. [17] proposed a method for diversifying query results in online shopping applications wherein the query is presented in a structure form using online forms.

Our work is based on axiomatizations of ranking and clustering systems [2, 11, 3]. [11] proposed a set of three natural axioms for clustering functions and showed that no clustering function satisfies all three axioms. [3] study ranking functions that combine individual votes of agents into a social ranking of the agents and compare them to social choice welfare functions which were first proposed in the classical work on social choice theory by [2].

[1] See [15] and references therein.

We also show a mapping from diversification functions to those used in facility dispersion [13, 12]. The interested reader will find a useful literature in the chapter on facility dispersion in [14, 5].

2 Axiomatic Framework

This section introduces the axiomatic framework and fixes the notation to be used in the remainder of the paper. We are given a set $U = \{u_1, u_2, \ldots, u_n\}$ of $n \geq 2$ of documents, and a finite set of queries Q. Now, given a query $q \in Q$ and an integer k, we want to output a subset $S_k \subseteq U$ (where $|S_k| = k$) of documents that is simultaneously both relevant and diverse.[2] The relevance of each document is specified by a function $w : U \times Q \to \mathbf{R}^+$, where a higher value implies that the document is more relevant to a particular query. The diversification objective is intuitively thought of as giving preference to dissimilar documents. To formalize this, we define a distance function $d : U \times U \to \mathbf{R}^+$ between the documents, where smaller the distance, the more similar the two documents are. We also require the distance function to be discriminative, i.e. for any two documents $u, v \in U$, we have $d(u, v) = 0$ if and only if $u = v$, and symmetric, i.e $d(u, v) = d(v, u)$. Note that the distance function need not be a metric.

Formally, the set selection function $f : 2^U \times Q \times w \times d \to \mathbf{R}$ can be thought of as assigning scores to all possible subsets of U, given a query $q \in Q$, a weight function $w(\cdot)$, a distance function $d(\cdot, \cdot)$. Fixing q, $w(\cdot)$, $d(\cdot, \cdot)$ and a given integer $k \in \mathbf{Z}^+$ ($k \geq 2$), the objective is to select a set $S_k \subseteq U$ of documents such that the value of the function f is maximized, i.e. the objective is to find

$$S_k^* = \operatorname*{argmax}_{S_k \subseteq U,\ |S_k| = k} f(S_k, q, w(\cdot), d(\cdot, \cdot)) \tag{1}$$

where all arguments other than the set S_k are fixed inputs to the function.

An important observation is that the diversification framework is under-specified and even if one assumes that the relevance and distance functions are provided, there are many possible choices of the diversification objective function f. These functions could trade-off relevance and similarity in different ways, and one needs to specify criteria for selection among these functions. A natural mathematical approach in such a situation is to provide axioms that any diversification system should be expected to satisfy and therefore provide *some* basis of comparison between different diversification functions.

2.1 Axioms of diversification

We propose that f is such that it satisfy the set of axioms given below, each of which is a property that is intuitive for the purpose of diversification. In addition, we show that any proper subset of these axioms is *maximal*, i.e. no diversification function can satisfy all these axioms. This provides a natural method of selecting between various objective functions, as one can choose the essential properties for any particular diversification system. In section 3, we will illustrate the use of the axioms in choosing between different diversification objectives. Before we state the axioms, we state the following notation. Fix any q, $w(\cdot)$, $d(\cdot, \cdot)$, k and f, such that f is maximized by S_k^*, i.e., $S_k^* = \operatorname{argmax}_{S_k \subseteq U} f(S_k, q, w(\cdot), d(\cdot, \cdot))$.

1. **Scale invariance**: Informally, this property states that the set selection function should be insensitive to the scaling of the input functions. Consider the optimal set S_k^*. Now, we require f to be such that we have $S_k^* = \operatorname{argmax}_{S_k \subseteq U} f(S_k, q, \alpha \cdot w(\cdot), \alpha \cdot d(\cdot, \cdot))$ for any fixed positive constant $\alpha \in \mathbf{R}$, $\alpha > 0$, i.e. S_k^* still maximizes f even if all relevance and distance values are scaled by some constant. Note that we require

[2] In this work, we focus our attention on the *set selection* problem instead of producing a *ranked list* as the output, which is the ultimate goal in some settings such as web search. This choice is motivated by the fact that the set selection problem captures the core trade-off involved in building a diversification system as we will see shortly. Furthermore, there are several ways to convert the selected set into a ranked list. For instance, one can always rank the results in order of relevance.

the constant to be the same for both $w(\cdot)$ and $d(\cdot,\cdot)$ in order to be consistent with respect to the scale of the problem.

2. **Consistency**: Consistency states that making the output documents more relevant and more diverse, and making other documents less relevant and less diverse should not change the output of the ranking. Now, given any two functions $\alpha : U \to \mathbf{R}^+$ and $\beta : U \times U \to \mathbf{R}^+$, we modify the relevance and weight functions as follows:

$$w(u) = \begin{cases} w(u) + \alpha(u) & , u \in S_k^* \\ w(u) - \alpha(u) & , \text{otherwise} \end{cases}$$

$$d(u,v) = \begin{cases} d(u,v) + \beta(u,v) & , u, v \in S_k^* \\ d(u,v) - \beta(u,v) & , \text{otherwise} \end{cases}$$

The ranking function f must be such that it is still maximized by S_k^*. We emphasize that the change in the relevance and diversity for each document can be different, and consistency only requires the function f to be invariant with respect to the right *direction* of the change.

3. **Richness**: Informally speaking, the richness condition states that we should be able to achieve any possible set as the output, given the right choice of relevance and distance function. This property is motivated by the practical fact that one could construct a query (along with corresponding relevance and diversity functions) that would correspond to any given output set. Formally, there exists some $w(\cdot)$ and $d(\cdot,\cdot)$ such that for any $k \geq 2$, there is a unique S_k^* which maximizes f.

4. **Stability**: The stability condition seeks to ensure that the output set does not change arbitrarily with the output size, i.e., the function f should be defined such that $S_k^* \subset S_{k+1}^*$. One can also consider relaxations of the strict stability condition such as $|S_k^* \cap S_{k+1}^*| > 0$. We do not discuss the relaxation further in this work.

5. **Independence of Irrelevant Attributes**: This axiom states that the score of a set is not affected by most attributes of documents outside the set. Specifically, given a set S, we require the function f to be such that $f(S)$ is independent of values of both $w(u)$ and $d(u,v)$ for all $u, v \notin S$.

6. **Monotonicity**: Monotonicity simply states that the addition of any document does not decrease the score of the set. Fix any $w(\cdot)$, $d(\cdot,\cdot)$, f and $S \subseteq U$. Now, for any $x \notin S$, we must have $f(S \cup \{x\}) \geq f(S)$.

Finally, we state two properties that ensure that the diversification system will not trivially ignore one of either relevance or diversity objectives. This ensures that the system will capture the trade-off between the two objectives in a non-degenerate manner.

7. **Strength of Relevance**: This property ensures that no function f ignores the relevance function. Formally, we fix some $w(\cdot)$, $d(\cdot,\cdot)$, f and S. Now, the following properties should hold for any $x \in S$:

   (a) There exist some real numbers $\delta_0 > 0$ and $a_0 > 0$, such that the condition stated below is satisfied after the following modification: obtain a new relevance function $w'(\cdot)$ from $w(\cdot)$, where $w'(\cdot)$ is identical to $w(\cdot)$ except that $w'(x) = a_0 > w(x)$. The remaining relevance and distance values could decrease arbitrarily. Now, we must have

$$f(S, w'(\cdot), d(\cdot,\cdot), k) = f(S, w(\cdot), d(\cdot,\cdot), k) + \delta_0$$

(b) If $f(S \setminus \{x\}) < f(S)$, then there exist some real numbers $\delta_1 > 0$ and $a_1 > 0$ such that the following condition holds: modify the relevance function $w(\cdot)$ to get a new relevance function $w'(\cdot)$ which is identical to $w(\cdot)$ except that $w'(x) = a_1 < w(x)$. Now, we must have

$$f(S, w'(\cdot), d(\cdot,\cdot), k) = f(S, w(\cdot), d(\cdot,\cdot), k) - \delta_1$$

8. **Strength of Similarity**: This property ensures that no function f ignores the similarity function. Formally, we fix some $w(\cdot)$, $d(\cdot,\cdot)$, f and S. Now, the following properties should hold for any $x \in S$:

(a) There exist some real numbers $\delta_0 > 0$ and $b_0 > 0$, such that the condition stated below is satisfied after the following modification: obtain a new distance function $d'(\cdot,\cdot)$ from $d(\cdot,\cdot)$, where we increase $d(x,u)$ for the required $u \in S$ to ensure that $\min_{u \in S} d(x,u) = b_0$. The remaining relevance and distance values could decrease arbitrarily. Now, we must have

$$f(S, w(\cdot), d'(\cdot,\cdot), k) = f(S, w(\cdot), d(\cdot,\cdot), k) + \delta_0$$

(b) If $f(S \setminus \{x\}) < f(S)$, then there exist some real numbers $\delta_1 > 0$ and $b_1 > 0$ such that the following condition holds: modify the distance function $d(\cdot,\cdot)$ by decreasing $d(x,u)$ to ensure that $\max_{u \in S} d(x,u) = b_1$. Call this modified distance function $d'(\cdot,\cdot)$. Now, we must have

$$f(S, w(\cdot), d'(\cdot,\cdot), k) = f(S, w(\cdot), d(\cdot,\cdot), k) - \delta_1$$

Given these axioms, a natural question is to characterize the set of functions f that satisfy these axioms. A somewhat surprising observation here is that it is impossible to satisfy all of these axioms simultaneously (proof is in the full paper [8]):

Theorem 1: No function f satisfies all 8 axioms stated above.

This result allows us to naturally characterize the set of diversification functions, and selection of a particular function reduces to deciding upon the subset of axioms (or properties) that the function is desired to satisfy. The next section explore this idea further and shows that the axiomatic framework could be a powerful tool in choosing between diversification function. Another advantage of the framework is that it allows a theoretical characterization of the function which is independent of the specifics of the diversification system such as the distance and the relevance function.

3 Objectives and Algorithms

In light of the impossibility result shown in Theorem 1, we can only hope for diversification functions that satisfy a subset of the axioms. We note that the list of such functions is possibly quite large, and indeed several such functions have been previously explored in the literature (see [6, 16, 1], for instance). Further, proposing a diversification objective may not be useful in itself unless one can actually find algorithms to optimize the objective. In this section, we aim to address both of the above issues: we demonstrate the power of the axiomatic framework in choosing objectives, and also propose reductions from a number of natural diversification objectives to the well-studied combinatorial optimization problem of facility dispersion [14]. In particular, we propose two diversification objectives in the following sections, and provide algorithms that optimize these objectives. We also present a brief characterization of the objective functions studied in earlier works [1, 16, 6]. We will use the same notation as in the previous section and have the objective (namely equation 1), where f would vary from one function to another. Also, we assume $w(\cdot)$, $d(\cdot,\cdot)$ and k to be fixed here and hence use the shorthand $f(S)$ for the function.

3.1 Max-sum diversification

A natural bi-criteria objective is to maximize the sum of the relevance and dissimilarity of the selected set. This objective can be encoded in terms of our formulation in terms of the function $f(S)$, which is defined as follows:

$$f(S) = (k-1)\sum_{u\in S} w(u) + 2\lambda \sum_{u,v\in S} d(u,v) \tag{2}$$

where $|S| = k$, and $\lambda > 0$ is a parameter specifying the trade-off between relevance and similarity. Observe that we need to scale up the first sum to balance out the fact that there are $\frac{k(k-1)}{2}$ numbers in the similarity sum, as opposed to k numbers in the relevance sum. We first characterize the objective in terms of the axioms.

Remark 1: The objective function given in equation 2 satisfies all the axioms, except stability.

This objective can be recast in terms of a facility dispersion objective, known as the MaxSumDispersion problem. The MaxSumDispersion problem is a facility dispersion problem having the objective maximizing the sum of all pairwise distances between points in the set S which we show to be equivalent to equation 2. To this end, we define a new distance function $d'(u,v)$ as follows:

$$d'(u,v) = w(u) + w(v) + 2\lambda d(u,v) \tag{3}$$

It is not hard to see the following claim (proof skipped):

Claim 2: $d'(\cdot,\cdot)$ is a metric if the distance $d(\cdot,\cdot)$ constitutes a metric.

Further, note that for some $S \subseteq U$ ($|S| = k$), we have:

$$\sum_{u,v\in S} d'(u,v) = (k-1)\sum_{u\in S} w(u) + 2\lambda \sum_{u,v\in S} d(u,v)$$

using the definition of $d'(u,v)$ and the fact that each $w(u)$ is counted exactly $k-1$ times in the sum (as we consider the complete graph on S). Hence, from equation 2 we have that

$$f(S) = \sum_{u,v\in S} d'(u,v)$$

But this is also the objective of the MaxSumDispersion problem described above where the distance metric is given by $d'(\cdot,\cdot)$.

Given this reduction, we can map known results about MaxSumDispersion to the diversification objective. First of all, we observe that maximizing the objective in equation 2 is NP-hard, but there are known approximation algorithms for the problem. In particular, there is a 2-approximation algorithm for the MaxSumDispersion problem [10, 9] (for the metric case) and is given in algorithm 1. Hence, we can use algorithm 1 for the max-sum objective stated in 2.

3.2 Mono-objective formulation

The space of diversification objectives is quite rich, and indeed there are several examples of objectives that do not reduce to facility dispersion. For instance, the second objective we will explore does not relate to facility dispersion as it combines the relevance and the similarity values into a single value for each *document* (as opposed to each edge for the previous two objectives). The objective can be stated in the notation of our framework in terms of the function $f(S)$, which is defined as follows:

$$f(S) = \sum_{u\in S} w'(u) \tag{4}$$

Algorithm 1 Algorithm for MAXSUMDISPERSION

Require: Universe U, k

Ensure: Set S ($|S| = k$) that maximizes $f(S)$

Initialize the set $S = \emptyset$
for $i \leftarrow 1$ **to** $\lfloor \frac{k}{2} \rfloor$ **do**
    Find $(u, v) = \text{argmax}_{x,y \in U}\, d(x, y)$
    Set $S = S \cup \{u, v\}$
    Delete all edges from E that are incident to u or v
end for
if k is odd **then**
    add an arbitrary document to S
end if

where the new relevance value $w'(\cdot)$ for each document $u \in U$ is computed as follows:

$$w'(u) = w(u) + \frac{\lambda}{|U| - 1} \sum_{v \in U} d(u, v)$$

for some parameter $\lambda > 0$ specifying the trade-off between relevance and similarity. Intuitively, the value $w'(u)$ computes the "global" importance (i.e. not with respect to any particular set S) of each document u. The axiomatic characterization of this objective is as follows:

Remark 2: The objective in equation 4 satisfies all the axioms except consistency.

Also observe that it is possible to exactly optimize objective 4 by computing the value $w'(u)$ for all $u \in U$ and then picking the documents with the top k values of u for the set S of size k.

3.3 Other objective functions

We note that the link to the facility dispersion problem explored in section 3.1 is particularly rich as many dispersion objectives have been studied in the literature (see [14, 5]). Although we only explored a single objective here in order to illustrate the use of the dispersion objective, similar reductions can be used to obtain algorithms with provable guarantees for many other diversification objectives [8].

The axiomatic framework can also be used to characterize diversification objectives that have been proposed previously. For instance, we note that the DIVERSIFYobjective function in [1] as well as the MINQUERY-ABANDONMENT formulations proposed in [16] violate the stability and the independence of irrelevant attributes axioms.

4 Conclusions

This work presents an approach to characterizing diversification systems using a set of natural axioms. The choice of axioms presents an objective basis for characterizing diversification objectives independent of the algorithms used, and the specific forms of the distance and relevance functions. Specifically, we illustrate the use of the axiomatic framework by studying two objectives satisfying different subsets of axioms.

References

[1] R. Agrawal, S. Gollapudi, A. Halverson, S. Ieong, *Diversifying Search Results*, in Proc. 2nd ACM Intl Conf on Web Search and Data Mining, 5–14, 2009.

[2] K. Arrow, *Social Choice and Individual Values*, Wiley, New York, 1951.

[3] A. Altman, M. Tennenholtz, *On the axiomatic foundations of ranking systems*, in Proc. 19th Int'l Joint Conference on Artificial Intelligence, 917–922, 2005.

[4] J. Carbonell, J. Goldstein, *The use of MMR, diversity-based reranking for reordering documents and producing summaries*, in Proc. of the Int'l Conf. on Research and Development in Information Retrieval (SIGIR), 335–336, 1998.

[5] B. Chandra, M. M. Halldórsson, *Approximation Algorithms for Dispersion Problems*, in J. Algorithms, (38):2, 438–465, 2001.

[6] H. Chen, D. R. Karger, *Less is more: probabilistic models for retrieving fewer relevant documents*, in Proc. of the Int'l Conf. on Research and Development in Information Retrieval (SIGIR), 429–436, 2006.

[7] C. L. A. Clarke, M. Kolla, G. V. Cormack, O. Vechtomova, A. Ashkan, S. Büttcher, I. MacKinnon, *Novelty and diversity in information retrieval evaluation*, in Proc. of the Int'l Conf. on Research and Development in Information Retrieval (SIGIR), 659–666, 2008.

[8] S. Gollapudi, A. Sharma, *An Axiomatic Approach for Result Diversification*, in Proc. of the Int'l Conf. on World Wide Web, 381–390, 2009.

[9] R. Hassin, S. Rubinstein, A. Tamir, *Approximation algorithms for maximum dispersion*, in Operations Research Letters, (21):3, 133–137, 1997.

[10] B. Korte, D. Hausmann, *An Analysis of the Greedy Heuristic for Independence Systems*, in Algorithmic Aspects of Combinatorics, (2), 65–74, 1978.

[11] J. Kleinberg, *An Impossibility Theorem for Clustering*, in Advances in Neural Information Processing Systems 15: Proc. of the 2002 Conference, MIT Press, 2003.

[12] S. S. Ravi, D. J. Rosenkrantz, G. K. Tayi, *Facility dispersion problems: Heuristics and special cases*, in Proc. 2nd Workshop on Algorithms and Data Structures (WADS), 355–366, 1991.

[13] S. S. Ravi, D. J. Rosenkrantz, G. K. Tayi, *Heuristic and special case algorithms for dispersion problems*, in Operations Research, (42):2, 299–310, 1994.

[14] S. S. Ravi, D. J. Rosenkrantzt, G. K. Tayi, *Approximation Algorithms for Facility Dispersion*, in Handbook of Approximation Algorithms and Metaheuristics, Chapman & Hall/CRC, 2007.

[15] S. Robertson, H. Zaragoza, *On rank-based effectiveness measures and optimization*, in Inf. Retr. (10):3, 321–339, 2007.

[16] A. D. Sarma, S. Gollapudi, S. Ieong, *Bypass rates: reducing query abandonment using negative inferences*, in Proc. of the ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining, 177–185, 2008.

[17] E. Vee, U. Srivastava, J. Shanmugasundaram, P. Bhat, S. A. Yahia, *Efficient Computation of Diverse Query Results*, in IEEE 24th Int'l Conference on Data Engineering (ICDE), 228–236, 2008.

[18] C. X. Zhai, W. W. Cohen, J. Lafferty, *Beyond independent relevance: methods and evaluation metrics for subtopic retrieval*, in Proc. of the Int'l Conf. on Research and Development in Information Retrieval (SIGIR), 10–17, 2003.

[19] C. X. Zhai, *Risk Minimization and Language Modeling in Information Retrieval*, in Carnegie Mellon University, 2002.

[20] C. X. Zhai and J. Lafferty, *A risk minimization framework for information retrieval*, in Information Processing and Management, (42):1, 31–55, 2006.

[21] C. N. Ziegler, S. M. McNee, J. A. Konstan, G. Lausen, *Improving recommendation lists through topic diversification*, in Proc. of the Int'l Conf. on World Wide Web, 22–32, 2005.