Parameter Tuning of Reranking-based Diversification Algorithms using Total Curvature Analysis

Shameem Puthiya Parambath
sham.puthiya@glasgow.ac.uk
University of Glasgow
Glasgow, United Kingdom

Siwei Liu
s.liu.4@research.gla.ac.uk
University of Glasgow
Glasgow, United Kingdom

Christos Anagnostopoulos
christos.anagnostopoulos@glasgow.ac.uk
University of Glasgow
Glasgow, United Kingdom

Roderick Murray-Smith
roderick.murray-smith@glasgow.ac.uk
University of Glasgow
Glasgow, United Kingdom

Iadh Ounis
iadh.ounis@glasgow.ac.uk
University of Glasgow
Glasgow, United Kingdom

ABSTRACT

In this paper, we analyze re-ranking based recommendation diversification algorithms and observe that, commonly, such algorithms can be unified under the scheme of maximizing submodular or modular objective functions from the class of parameterized concave over modular functions. We showcase that such diversification objective functions can be expressed in a generic functional form consisting of the relevance and diversity terms. We then theoretically analyze and show that the *total curvature* of submodular functions provides insights about the relevance-diversity trade off. This is expected to support data analysts to seek balanced hyper-parameter values and, thus, serve as a 'vehicle of validation' by checking the total curvature of submodular objective functions. Our experimental evaluation and performance assessment over benchmark datasets are aligned with our theoretical analysis. We also discuss the importance of balanced trade-off between relevance and diversity in specific application settings like news recommendations to trade-off algorithmic bias and short term user engagement.

CCS CONCEPTS

• **Information systems → Recommender systems**; *Personalization*; **Information retrieval diversity**.

KEYWORDS

recommender systems, relevance-diversity trade-off, submodular functions

ACM Reference Format:
Shameem Puthiya Parambath, Siwei Liu, Christos Anagnostopoulos, Roderick Murray-Smith, and Iadh Ounis. 2022. Parameter Tuning of Reranking-based Diversification Algorithms using Total Curvature Analysis. In *Proceedings of the 2022 ACM SIGIR International Conference on the Theory of Information Retrieval (ICTIR '22), July 11–12, 2022, Madrid, Spain.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3539813.3545135

ICTIR '22, July 11–12, 2022, Madrid, Spain

ACM ISBN 978-1-4503-9412-3/22/07…$15.00
https://doi.org/10.1145/3539813.3545135

1 INTRODUCTION

In prior work, most of the standard personalized recommendation algorithms consider only the relevance of the items to users upon proceeding with predictions [1, 5, 14]. This relevance-driven strategy implicitly assumes that the degrees of relevance of the items within the recommended set are independent of each other. Undeniably, this results in similar and redundant items to appear in the recommended set [7]. Moreover, popularity bias [1, 31] in Recommender Systems (RS) exacerbates the situation by excluding unpopular, but relevant items from recommendations though. In personalized recommendation, diversification is established as the vehicle to recommend novel, serendipitous items that result in higher user satisfaction [2, 26, 27, 35, 40]. Furthermore, in the context of group recommendation, diversification is a way to generate consensus recommendations by finding items relevant to a group of users [25]. Additionally, diversification is a tool to handle algorithmic bias of a predictive model. It is already established that the majority of the data is generated by a small set of users and follows Zipf's law, also known as the wisdom of a few [4]. Hence, there is an inherent data bias and any learning algorithm trained on such data will inherit such a bias. Diversification approaches deal with algorithmic bias by countering popularity bias [31].

A common strategy for diversification is re-ranking, i.e. given a list of recommended items, re-ranking algorithms reorder the list to account for diversity [14]. Many re-ranking based diversification algorithms have been proposed in the past, which exploit different aspects of recommendations, e.g., personal popularity tendency [22], genre coverage [35], interest coverage [27], and long-tail recommendation [2]. It has also been shown that re-ranking algorithms make use of the 'diminishing return' property of submodular functions to trade-off between relevance and diversity [9].

Submodular functions are set functions with non-increasing marginal gains, i.e. by adding an element to a *subset* of a given set $\mathcal{A}$, it yields at least much value (or more) as if we add the element to the *whole* $\mathcal{A}$ itself. Intuitively, given a set, adding a new item similar to some already existing items yields lower value than adding an item dissimilar to some already existing items (refer to Appendix A for a formal definition).

Even though re-ranking based diversification algorithms are very popular, the objective functions used for diversification are poorly understood. Additionally, balanced relevance-diversity trade-off is obtained by hyperparameters tuning tasks like grid search. In this paper, we aim to: (i) provide a unified theoretical framework for different re-ranking based diversification algorithms; and (ii) propose a method to validate the balanced relevance-diversity trade-off parameter by analyzing the *total curvature* of the diversification objective functions. Our analysis reveals that re-ranking objective functions are not only submodular (or modular). Instead, they take a simple form: the composition of a concave (including linear) function with a modular one. We illustrate that the majority of the diversification objective functions can be written in our proposed, generic, yet simple form consisting of relevance and diversity terms. Such a generic representation will help data analysts to design new application/data specific diversification algorithms and tune the hyperparameter to achive balanced trade-off. We also show that analyzing the total curvature of such objective functions provides meaningful insights about relevance-diversity trade-off and then one can adjust the trade-off parameter accordingly. To this end, the total curvature value will serve as a 'vehicle of validation' of the balanced trade-off hyperparameter setting. It is worth noting that the most favourable relevance-diversity trade-off may not be the balanced trade-off in production, as it depends on business objectives (like always recommending a serendipitous item or include a sponsored item), diversity and/or relevance constraints (like showing minimum of 90% relevant items). Moreover, there is no single optimal point, but a Pareto frontier of relevance-diversity values [13]. However, the value of *total curvature* provides an indication of which direction the hyperparameter values should be adjusted depending on the intended objective.

The remainder of this paper is structured as follows. We give a brief overview of existing diversification algorithms and position our contributions in section 2. In section 3, we show that popular diversification algorithms are based on the class of parameterized concave/linear composition of modular functions. We also discuss the *total curvature* and the worst-case lower bound for submodular maximization in terms of the *total curvature*. Finally, we demonstrate our experimental results in section 5 before concluding the paper in section 6.

2 RELATED WORK

The recommendation diversification issue has been extensively studied in the past [3, 10, 14, 19, 22, 23, 27, 35, 36]. Many of these algorithms were inspired by Web search result diversification algorithms. Recommendation diversification algorithms can be classified into three schemes: (i) re-ranking based; (ii) multi-objective optimization based; (iii) multi-armed bandit based algorithms.

Given a recommendation list produced by any standard personalized recommendation algorithm (like user-user or item-item collaborative filtering), re-ranking algorithms re-order the items of this list such that relevant and diverse items appear in the top-k rankings. Re-ranking algorithms are simple to implement and widely popular since they can be readily plugged in to existing personalized RSs in a post-processing step. We will elaborate on an in-depth analysis of re-ranking based algorithms in section 3.

Multi-objective optimization based algorithms train a prediction model adopting joint optimization over an objective function taking into account the expected cost of relevance and diversity terms. Such algorithms extend the traditional collaborative filtering models by adding diversity in the objective function. The authors in [14] proposed an objective, which combines the standard latent factor model with intra-list distance as a measure of diversity. The authors in [10] considered the diversified recommendation as a structured supervised learning problem and proposed a structured Support Vector Machine-based objective function. The work in [11] proposed an explainable matrix factorization by incorporating a novelty term. Such approaches always involve solving complex optimization problems, often non-convex and, thus, resulting in extensively high computational cost while lacking of scalability.

Multi-armed bandit based algorithms are mainly applied on sequential recommendation tasks. In particular, the authors in [24, 39] proposed an Upper Confidence Bound diversification algorithm assuming submodular reward function, which was linear in features. Finally, the work in [28] proposed an algorithm for general reward function. The interested reader could be referred to a comprehensive survey of diversification algorithms in [8, 17, 38].

3 RERANKING-BASED DIVERSIFICATION FUNCTIONS

We denote the set of items using $\mathcal{X}$, the set of observed (rated) items using $\mathcal{O}$, the set of unobserved items using $\mathcal{E}$ and set of users by $\mathcal{U}$. $\mathcal{S}$ denotes a subset of $\mathcal{E}$ such that $|\mathcal{S}| \leq k$ for any positive k; $|\cdot|$ denotes set cardinality and k is the number of recommendations to be retrieved. Given a user $u \in \mathcal{U}$ and an item $i \in \mathcal{X}$, we denote the relevance of the item i for the user u as $rel_u(i) \in \mathbb{R}^+$. For a given set $\mathcal{S}$, $rel_u(\mathcal{S})$ represents the relevance of the items in the set $\mathcal{S}$ for user u. It is defined as the sum of the relevance of individual items in $\mathcal{S}$, i.e., $rel_u(\mathcal{S}) = \sum_{i \in \mathcal{S}} rel_u(i)$. We omit the subscript u from the forthcoming discussion wherever it is clear from the context.

3.1 Generic Functional Form

We start with a generic functional form for the re-ranking based diversification objective function and show that most of the previously proposed approaches can be expressed using this form. Given a set $\mathcal{S}$, consider the functional form given below:

$$F(\mathcal{S}) = f(\mathcal{S}) + \beta g(h(\mathcal{S})), \tag{1}$$

where $f(\cdot)$ is a modular function representing the relevance of the recommendation set $\mathcal{S}$ and $g(h(\mathcal{S}))$ represents the diversity term which is the composition of a linear or concave function, $g(\cdot)$, and a modular function h.[1] The β is the hyperparameter to be tuned for the relevance-diversity trade-off. Often, the $h(\cdot)$ function is defined as a function of relevance itself but capturing a certain diversity aspect. In typical settings, the notion of diversity is defined using different concepts, e.g., item coverage, popularity bias, item novelty, serendipity, and long-tail recommendation. We will show that the majority of the re-ranking based diversification objective functions are represented using our generic form in Equation (1) and only differ in the way $h(\cdot)$ and $g(\cdot)$ are defined.

[1]The functional form $\lambda f(\mathcal{S}) + (1-\lambda) g(h(\mathcal{S}))$ is equivalent to our form (normalized version) $f(\mathcal{S}) + \beta g(h(\mathcal{S}))$.

The TANGENT algorithm [23] proposed an objective function that can be written as $F(\mathcal{S}) = rel(S) - \beta \frac{1}{rel(S)}$ [2]. Here, $f(\cdot)$ is the relevance of the set of items, which is defined as the sum of relevance of items in $\mathcal{S}$ and, thus, a modular function of the set $\mathcal{S}$; while $g(\cdot)$ is the reciprocal of the relevance. Since the reciprocal function is convex for any positive values of $rel(\mathcal{S})$, g is concave (−ve of the reciprocal convex function) over the modular function $rel(\mathcal{S})$. It is to be noted that, here, the notion of diversity is defined as the reciprocal of relevance (higher relevance value results in lower reciprocals).

The diversity objective function proposed in [22] took the form of $F(\mathcal{S}) = rel(\mathcal{S}) + \log\left(\frac{h(\mathcal{S})}{h(O)}\right)$. The diversity notion, based on the popularity bias, is used and $h(\mathcal{S})$ is defined as the *personal popularity tendency* of the set $\mathcal{S}$ with respect to a given user. The aim of this diversity objective is to encourage recommending items from the item tail distribution. Moreover, the diversity term is defined as the composition of a concave function (log) with the modular function $\frac{h(\mathcal{S})}{h(O)}$.

The algorithm BinomDiv[35] re-ranks the recommended items by defining a diversity term in terms of coverage. Coverage-based diversity objectives encourage items from different parts of the item distribution to be ranked in top positions. For instance, topic coverage or genre coverage is a popular diversity metric, which discourages items from the same topics to be ranked together in top positions. Similar to the aforementioned cases, the $f(\cdot)$ function in BinomDiv corresponds to a modular function indicating the relevance of an item, while the diversity term is defined using coverage term and non-redundancy. The exact objective function takes the form $(1-\beta) \cdot rel(\mathcal{S}) + \beta \cdot h(\mathcal{S})$ The function $h(\mathcal{S})$ defined as the product of a coverage term and a non-redundancy term. Both, coverage and non-redundancy terms, are defined in terms of the topic coverage and, thus, take the form x^{β} with $\beta \in (0, 1)$. Here, x is the probability that a relevant topic is not recommended by a recommendation set. Since x^{β} is concave for any $\beta \in (0, 1)$, the diversity term is the composition of a concave over modular function.

The objective function proposed in [37] consists of two terms: the first modular term captures the relevance with respect to neighbouring users, while the second term, the neighbour coverage function, stands as a surrogate for diversity. The diversity term is defined as the concave composition $g(x) = \frac{x}{1+x}$ of a positive modular function.

Steck [33] used Maximal Margin Relevance (MMR) inspired objective function defined in [6] to calibrate the recommendation. Though calibration can be different from diversity, the calibration technique indirectly induces diversity by maximizing users' interest coverage [33]. Unlike in other MMR based approaches, Steck [33] used a submodular diversity term defined in terms of the concave function, log. Moreover, Abdollahpouri et al. [2] proposed an algorithm for recommending items from long-tail distribution. The core idea is based on *xQuAD* algorithm proposed in [29] which is based on MMR heuristic, and hence submodular.

It is also very common that only the diversity aspect is considered for re-ranking, i.e., $f(\mathcal{S}) = c$ for a constant c, in Equation (1). The authors in [34] proposed a re-ranking algorithm by disregarding the relevance aspect and considering only the topic-coverage of the recommended items. The objective function considers only the diversity aspect and thus takes the form

$$F(\mathcal{S}) = \sum_{i \in C(\mathcal{S})} \sum_{j \in C(\mathcal{S}_{-i})} \frac{-1}{|C(\mathcal{S})||C(\mathcal{S}_{-i})|},$$

where $C(\mathcal{S})$ represents the topics associated with the set of items $\mathcal{S}$ and $\mathcal{S}_{-i} = \mathcal{S} \setminus \{i\}$. Here, $g(\cdot)$ is the sum of the concave functions $-1/x$. Puthiya Parambath et al. [27] also proposed a re-ranking algorithm considering only the diversity term. In fact, the proposed objective function combines the relevance and diversity objectives in a single objective defined in terms of the relevant interest coverage. They proposed different concave compositions of the interest coverage term x like x^{β}, $\beta \in [0, 1]$ and $\log x$ to re-rank the items. In addition, the proposed method achieves diversification in a single stage unlike other re-ranking algorithms. In [36], the authors proposed an intent-aware diversification algorithm. [9] showed that intent-aware objective functions are either submodular or modular functions. The objective function in [36] is modular since it is based on re-weighting item based collaborative filtering scores using intent-aware covariance values. Similar to [27], this approach also achieves diversification in a single step.

In the seminal work by Carbonell and Goldstein [6], the authors introduced Maximal Margin Relevance (MMR), a method to re-rank web search results to induce the diversity. They employed the following heuristic for adding an element i to the set $\mathcal{S}$:

$$\beta \cdot rel(i) - (1-\beta) \max_{j \in \mathcal{S}} sim(i, j), \qquad (2)$$

where $rel(i)$ and $\max_{j \in \mathcal{S}} sim(i, j)$ are positive modular functions of $\mathcal{S}$. $sim(i, j)$ is the positive similarity function between items i and j. Here, $-sim(i, j)$ acts as the diversity term because lower value of similarity for an item will encourage that item to be included in the final recommendation list as per Equation 2. The β factor in the above formulation is the hyperparameter that is tuned for relevance-diversity trade-off. The MMR heuristic is defined as the difference between two modular functions, i.e., a linear composition of modular functions and it is shown to be a submodular function, albeit not necessarily monotonic [18].

Many MMR inspired objectives form the basis for different diversification algorithms [5, 16, 20] and they differ in the way relevance and diversity terms are defined. For example, in [20], authors defined diversity in terms of personality trait of a user. However, in [5] diversity term is replaced with relative diversity which is defined as the sum of the dissimilarity between the already recommended items and a new, to be recommended, item. [16] defined diversity in terms of sub-profile coverage of items for each user. Since MMR is defined as a greedy heuristic [6], the exact objective function used in MMR is not known. However, all the MMR inspired diversification algorithms discussed above can be re-written in our standard form as shown in Equation (1).

Table 1 summarizes the essence of the above discussion pertaining to re-ranking algorithms. We conclude this section by rephrasing the characteristic of the re-ranking algorithm objective function: *it is a submodular/modular function belonging to the class of concave/linear over modular functions.* Though such an observation

[2]In the original formulation the two terms are added with β assumed to be negative.

Algorithm/Method	$g(x)$
[23, 34]	$-\frac{1}{x},\ x > 0,$
[22]	$\log(x),\ x > 0$
[27, 35]	$x^{\beta},\ x \geq 0, \beta \in (0, 1)$
[37]	$\frac{x}{1+x},\ x \geq 0$
[36]	$\beta x,\ \beta > 0$

Table 1: Functional form of $g(x)$

might look straightforward, to the best of our knowledge, it has not been formally established in the literature. The positive side of this observation is that practitioners/data analysts can easily develop a re-ranking strategy for diversification based on the available data in hand, depending on the business objectives with their specific diversity definitions.

4 OPTIMALITY OF RE-RANKING DIVERSIFICATION ALGORITHMS

The major factor behind the popularity of the re-ranking based diversification algorithms is that a solution can be obtained using a simple greedy heuristic. Moreover, the celebrated result due to [21] states that steepest ascent greedy algorithm for monotone submodular maximization guarantees a constant approximation worst case lower bound, i.e., $F(\mathcal{S}^*) \geq (1 - 1/e)F(\mathcal{S}^{opt})$, where e is the base of the natural logarithm, $\mathcal{S}^*$ is the greedy solution and $\mathcal{S}^{opt}$ is the unknown optimal solution. In practice, the greedy algorithm can often perform much better than this worst-case guarantee. The constant approximation guarantee of the greedy algorithm is applicable only when the objective function is monotone. In practice, this aspect is ignored. Though greedy algorithms for non-monotone submodular functions do not yield any theoretical guarantees regarding the optimality of the result, non-monotone submodular functions are commonly used for diversification purposes. The monotonicity property of different diversification algorithms is summarized in Table 2 in Appendix B.

4.1 Diversity-Relevance Trade-off Analysis using Total Curvature

We analyze the relevance-diversity trade-off in re-ranking algorithms to get a deeper theoretical understanding. In the majority of re-ranking algorithms, the trade-off in relevance and diversity is obtained by tuning the associated hyperparameter. For example, in the case of MMR [6] based algorithms, β is the hyperparameter controlling the trade-off. Similarly, in non-MMR based algorithms like [27, 35] also the hyperparameter β controls the relevance-diversity trade-off. Some formulations, particularly the modular formulations like in [3], do not provide the flexibility of explicit relevance-diversity trade-off. A typical question that arises in this context is: *what value of β gives balanced relevance-diversity trade-off?* We should note that for any given values of β, relevance-diversity trade-off obtained by re-ranking based diversification algorithm is Pareto optimal as it comes under the weighted-sum method [13]. However, it might not be a balanced one. A balanced trade-off will potentially depends on the application specific requirements.

The total curvature associated with submodular functions provides a way to find the balanced trade-off between relevance and diversity. The *total curvature* of a non-decreasing submodular set function with respect to a set $\mathcal{S}$ is defined as [12]:

$$\alpha = \max_{j \in \mathcal{S}} \frac{F(\mathcal{S} \setminus \{j\}) + F(\{j\}) - F(\mathcal{S})}{F(\{j\})}. \tag{3}$$

Intuitively, the *total curvature* measures how far $F(\cdot)$ is from being modular, and Equation 3 represents the distance of a monotone submodular function to the modularity. The total curvature can take values between 0 and 1 so that it is zero in the case of modular functions and, one in the case of matroid rank function [**?**].

The work in [12] extended the result of [21] and provided a tighter lower bound for the submodular maximization problem in terms of *total curvature*. According to [12], $F(\mathcal{S}^*) \geq {}^{(1-e^{-\alpha})F(\mathcal{S}^{opt})}/\alpha$. When the curvature is 1, this gives the standard approximation bound given in [21] and, for any other values of α, it strengthens the approximation lower bound in [21]. For example, if the curvature value $\alpha = 0.1$, we are guaranteed that $F(S^*) \geq 0.95F(S^{opt})$ compared to the standard result of $F(S^{opt}) \geq 0.63F(S^*)$. Given a set of items, the *total curvature* can be easily computed for any submodular function [15]. For a recommended set containing n items, *total curvature* can be calculated in $O(n)$. Here, we want to emphasize that curvature does not rely on specific functional form of $F(\cdot)$ but only on the marginal gains obtained by adding an item.

By changing the value of the hyperparameter β in the re-ranking objective function, one can effectively change the *total curvature* of the objective. Therefore, we can adjust the hyperparameter such that the re-ranking objective becomes a modular or submodular function with different curvature values. For example, by setting $\beta = 1$ in [6, 27, 35] or $\beta = 0$ in [23], the *total curvature* (α) of the resulting function can be made 0, thus, obtaining a modular function. Similarly, by adjusting the hyperparameter such that α becomes 1, one can obtain a submodular function close to a matroid rank function, which is the maximal distant from being modular. As α approaches 1, the objective changes from 'easy' to 'difficult'. A modular objective is considered 'easy' as the solution obtained using the greedy heuristic is optimal in contrast to that a submodular objective is 'difficult' as the greedy solution is sub-optimal.

Solving the modular objective function returns recommendation lists containing only relevant items with no diversity aspect. This also leads the greedy strategy to get the optimal objective value. To be specific, in the MMR case, if we take β such that $\alpha = 0$, only the relevant part remains, and the recommendation list contains items such that the intra-list dissimilarity is minimal. Similarly, if we adjust β in [27, 35] such that $\alpha = 1$, only the coverage part remains and the recommendation contains very diverse items at the expense of relevance. In other words, by simply adjusting the α to be zero or one, one forces the recommender system not to have any trade-off between relevance and diversity. For any $\alpha \in (0, 1)$, the greedy algorithm returns a mix of diverse and relevant items with the lower bound guarantee provided by the corresponding α value, and for α value closer to 0.5, we get the optimal balance between relevance and diversity.

Balanced Hyperparameter Tuning: In practical settings, the balanced value of β is found by the grid search and observing the performance on a validation set. Our above exposition suggests a more sophisticated way to perform a grid search. The balanced β

value depends on the objective function F, whereas α is independent of F [15]. An α value closer to 0.5 gives the balanced trade-off between being completely relevant and diverse. Thus, one can limit the grid search for β such that the corresponding α is closer to 0.5 and obtain the balanced trade-off in relevance and diversity. An α value closer to 0.5 indicates that the recommendation set contains a good mix of relevant and diverse items. Moreover, the value of α serves as a means of validation for the choice of the correct hyperparameter value. Therefore, we can justify whether the chosen hyperparameters are appropriate or not by directly observing the value of α. We empirically validate this claim in our experimental section.

5 EXPERIMENTAL EVALUATION

In this section, we aim to validate our claims on a movie recommendation task. The main goal of this section is two-fold: (i) to show that changing the hyperparameter β is *equivalent* to changing the *total curvature* of the diversification objective function; and (ii) to show that balanced trade-off in relevance-diversity can be obtained by choosing a β value corresponding to to the *total curvature* value close to 0.5. We used two diversification algorithms: (i) one based on *xQuAD*, which relies on MMR [2] and (ii) the coverage maximization algorithm in [27]. In case of *xQuAD*, we followed the *Binary xQuAD* algorithm as described in [2]. In case of [27], we experimented with two different concave compositions: (i) $g(x) = x^{\beta}$ and (ii) $g(x) = \log(x + \beta)$. Below, we discuss the results for $g(x) = x^{\beta}$.

We used the benchmark MovieLens 20M dataset. The rating values are block-wise ordered-taking one of the 5 values in {1,2,3,4,5}. For our experiments, we filter out all the users with less than 500 ratings and the final dataset contained 6,488,818 ratings for 7,322 users and 25,782 movies. We carried out holdout cross validation by splitting the rating data into training and test set such that 5% of the original data goes into testing and the remaining goes into training. The split is carried out five times in a manner that both the training and test sets span the entire user and movie sets. The reported results correspond to the average over the five splits. We used regularized weighted non-negative matrix factorization to extract the user and item features [32]. The extracted item features are used to estimate the item-item similarity matrix. As the regularizer, we used the variational form of the trace norm [30]. In case of [2], long-tail and short-head items are defined as mentioned in the paper. The values of α are calculated using the predicted recommendations for corresponding values of β as per Equation 3.

The performance of the recommendation task is evaluated on three metrics: one relevance ranking metric and two diversity metrics, where two diversity metrics include a serendipity metric and a metric to measure the distinctiveness of the recommendations. In our case, relevance is tied to the ranking, i.e., the most relevant item should be ranked in the top position. We use the popular Discounted Cumulative Gain (DCG) as the relevance ranking metric. DCG is a binary ranking metric and we discretized the observed rating value to calculate the DCG values on the test set. We used binary discretization such that rating values of 4 and 5 are deemed as relevant and as irrelevant, otherwise. We define Serendipity Score (SS) as the inverse of the average popularity of the recommended items, which are not rated by the user. We used Feature Distance (FD) as a measure of dissimilarity between the recommended items. It is defined as the average Euclidean distance between the item vectors in the recommended set and is similar to the intra-list distance [40]. SS and FD measure the diversity of the recommendations.

5.1 Experimental Results & Discussion

Due to the space limit figures are moved to Appendix C. The results of our experiments are shown in Figure 1 in Appendix C. The left side plots in Figure 1 show the trade-off between DCG and FD, while the right side plots show the trade-off between DCG and SS. The metrics DCG, FD, SS measure different properties and the orders of magnitude for these measures are different. The values reported in the plots are then normalized. We normalize each metric value by dividing the values with the maximum one reported for that specific metric (for different α, β values) to scale in (0,1]. The results show the metric values for top 5 recommendations. It is evident that as the hyperparameter β is changed from 0 to 1, the total curvature decreases from 1 to 0, i.e., the submodular function turns from 'difficult' to 'easy'. As β approaches 0, the diversification problem becomes modular function maximization and greedy algorithm returns the optimal solution containing only the most relevant items. In case of both algorithms, the relevance and diversity metrics intersect at α close to 0.5, thus, validating our theoretical analysis: an α value closer to 0.5 gives the balanced trade-off between being completely relevant and diverse. Thus, one can limit the grid search for β such that the corresponding α is closer to 0.5 and obtain the balanced trade-off in relevance and diversity. In practice, this fact serves as a validation for choosing the hyperparameter for balanced relevance-diversity trade-off. Nevertheless, the balanced trade-off may be different from the most favourable trade-off in production settings, which depends on e.g., intended business objectives, diversity or accuracy constraints. Yet, the value of *total curvature* gives an indication of which direction the hyperparameter value to be corrected depending on the business demands. We observed the same trend in our experiments using other concave formulations like $g(x) = \log(x + \beta)$ and $g(x) = 1 - e^{-\beta x}$.

6 CONCLUSIONS

We showed that re-ranking based diversification algorithms are based on maximizing a submodular objective function from the class of parameterized concave modular functions. Our analysis showed that studying the *total curvature* of submodular functions gives insights about the relevance-diversity trade-offs. We demonstrated that by varying the concave composition parameter, one effectively tunes the *total curvature* of the objective for the relevance-diversity trade-off.

ACKNOWLEDGEMENT

This work has been partially funded by the UK EPSRC grant #EP/R018634/1. Exploiting Closed-Loop Aspects in Computationally and Data Intensive Analytics.

REFERENCES

[1] Himan Abdollahpouri. 2019. Popularity bias in ranking and recommendation. In *AAAI Conference on Artificial Intelligence*. 529–530.

[2] Himan Abdollahpouri, Robin Burke, and Bamshad Mobasher. 2019. Managing popularity bias in recommender systems with personalized re-ranking. In *AAAI Conference on Artificial Intelligence*.

[3] Azin Ashkan, Branislav Kveton, Shlomo Berkovsky, and Zheng Wen. [n.d.]. Optimal greedy diversity for recommendation. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*.

[4] Ricardo Baeza-Yates. 2016. Data and algorithmic bias in the web. In *Proceedings of the 8th ACM Conference on Web Science*. 1–1.

[5] Keith Bradley and Barry Smyth. 2001. Improving recommendation diversity. In *Irish Conference on Artificial Intelligence and Cognitive Science*. 85–94.

[6] Jaime Carbonell and Jade Goldstein. 1998. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *SIGIR Conference on Research and development in information retrieval*. 335–336.

[7] Pablo Castells, Neil J Hurley, and Saul Vargas. 2015. Novelty and Diversity in Recommender Systems. *Recommender Systems Handbook* (2015), 881.

[8] J. Chakraborty and V. Verma. 2016. A survey of diversification techniques in recommendation systems. In *International Conference on Data Mining and Advanced Computing*. 35–40.

[9] Olivier Chapelle, Shihao Ji, Ciya Liao, Emre Velipasaoglu, Larry Lai, and Su-Lin Wu. 2011. Intent-based diversification of web search results: metrics and algorithms. *Information Retrieval* 14, 6 (2011), 572–592.

[10] Peizhe Cheng, Shuaiqiang Wang, Jun Ma, Jiankai Sun, and Hui Xiong. 2017. Learning to recommend accurate and diverse items. International World Wide Web Conference, 183–192.

[11] Ludovik Coba, Panagiotis Symeonidis, and Markus Zanker. 2019. Personalised novel and explainable matrix factorisation. *Data & Knowledge Engineering* 122 (2019), 142–158.

[12] Michele Conforti and Gérard Cornuéjols. 1984. Submodular set functions, matroids and the greedy algorithm: tight worst-case bounds and some generalizations of the Rado-Edmonds theorem. *Discrete applied mathematics* 7, 3 (1984), 251–274.

[13] M. Ehrgott and X. Gandibleux. 2002. *Multiple criteria optimization: state of the art annotated bibliographic surveys.* Kluwer Academic, Dordrecht.

[14] Neil J Hurley. 2013. Personalised ranking with diversity. In *ACM Conference on Recommender systems*. 379–382.

[15] Rishabh K Iyer, Stefanie Jegelka, and Jeff A Bilmes. 2013. Curvature and optimal algorithms for learning and minimizing submodular functions. In *Neural Information Processing Systems*. 2742–2750.

[16] Mesut Kaya and Derek Bridge. 2019. Subprofile-aware diversification of recommendations. *User Modeling and User-Adapted Interaction* (2019), 1–40.

[17] Matevž Kunaver and Tomaž Požrl. 2017. Diversity in recommender systems–A survey. *Knowledge-based systems* 123 (2017), 154–162.

[18] Hui Lin and Jeff Bilmes. 2011. A class of submodular functions for document summarization. In *Annual Meeting of the Association for Computational Linguistics*. 510–520.

[19] Siwei Liu, Iadh Ounis, Craig Macdonald, and Zaiqiao Meng. 2020. A heterogeneous graph neural model for cold-start recommendation. In *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval*. 2029–2032.

[20] Feng Lu and Nava Tintarev. 2018. A Diversity Adjusting Strategy with Personality for Music Recommendation.. In *Interfaces and Human Decision Making for Recommender Systems*. 7–14.

[21] George L Nemhauser, Laurence A Wolsey, and Marshall L Fisher. 1978. An analysis of approximations for maximizing submodular set functions. *Mathematical Programming* 14, 1 (1978).

[22] Jinoh Oh, Sun Park, Hwanjo Yu, Min Song, and Seung-T Park. 2011. Novel recommendation based on personal popularity tendency. In *International Conference on Data Mining*. 507–516.

[23] Kensuke Onuma, Hanghang Tong, and Christos Faloutsos. 2009. TANGENT: a novel,'Surprise me', recommendation algorithm. In *SIGKDD International Conference on Knowledge Discovery and Data Mining*. 657–666.

[24] Shameem Puthiya Parambath, Christos Anagnostopoulos, Roderick Murray-Smith, Sean MacAvaney, et al. 2021. Max-Utility Based Arm Selection Strategy For Sequential Query Recommendations. In *Asian Conference on Machine Learning*. PMLR, 564–579.

[25] Shameem Puthiya Parambath, Nishant Vijayakumar, and Sanjay Chawla. 2018. SAGA: A submodular greedy algorithm for group recommendation. In *AAAI Conference on Artificial Intelligence*.

[26] Shameem A Puthiya Parambath and Sanjay Chawla. 2020. Simple and effective neural-free soft-cluster embeddings for item cold-start recommendations. *Data Mining and Knowledge Discovery* 34, 5 (2020), 1560–1588.

[27] Shameem A Puthiya Parambath, Nicolas Usunier, and Yves Grandvalet. 2016. A coverage-based approach to recommendation diversity on similarity graph. In *ACM Conference on Recommender Systems*. 15–22.

[28] Lijing Qin, Shouyuan Chen, and Xiaoyan Zhu. 2014. Contextual combinatorial bandit and its application on diversified online recommendation. In *SIAM International Conference on Data Mining*. 461–469.

[29] Rodrygo LT Santos, Craig Macdonald, and Iadh Ounis. 2010. Exploiting query reformulations for web search result diversification. In *International World Wide Web Conference*. 881–890.

[30] Nathan Srebro, Jason Rennie, and Tommi S Jaakkola. 2005. Maximum-margin matrix factorization. In *Advances in neural information processing systems*. 1329–1336.

[31] Harald Steck. 2011. Item popularity and recommendation accuracy. In *ACM conference on Recommender systems*. 125–132.

[32] Harald Steck. 2013. Evaluation of recommendations: rating-prediction and ranking. In *ACM conference on Recommender systems*. 213–220.

[33] Harald Steck. 2018. Calibrated recommendations. In *ACM conference on Recommender systems*. 154–162.

[34] Ruilong Su, Li'Ang Yin, Kailong Chen, and Yong Yu. 2013. Set-oriented personalized ranking for diversified top-n recommendation. In *ACM conference on Recommender systems*. 415–418.

[35] Saúl Vargas, Linas Baltrunas, Alexandros Karatzoglou, and Pablo Castells. 2014. Coverage, redundancy and size-awareness in genre diversity for recommender systems. In *ACM conference on Recommender systems*. 209–216.

[36] Jacek Wasilewski and Neil Hurley. 2018. Intent-aware Item-based Collaborative Filtering for Personalised Diversification. In *Conference on User Modeling, Adaptation and Personalization*. 81–89.

[37] Le Wu, Qi Liu, Enhong Chen, Nicholas Jing Yuan, Guangming Guo, and Xing Xie. 2016. Relevance meets coverage: a unified framework to generate diversified recommendations. *ACM Trans. Intell. Syst. Technol.* 7, 3 (2016).

[38] Qiong Wu, Yong Liu, Chunyan Miao, Yin Zhao, Lu Guan, and Haihong Tang. 2019. Recent advances in diversified recommendation. *arXiv preprint arXiv:1905.06589* (2019).

[39] Yisong Yue and Carlos Guestrin. 2011. Linear submodular bandits and their application to diversified retrieval. In *Advances in Neural Information Processing Systems*. 2483–2491.

[40] Cai-Nicolas Ziegler, Sean M McNee, Joseph A Konstan, and Georg Lausen. 2005. Improving recommendation lists through topic diversification. In *International World Wide Web Conference*. 22–32.

A DEFINITIONS

Definition A.1. A function F defined on the subsets of a ground set $\mathcal{Z}$ is called submodular, if for all subsets $\mathcal{A}, \mathcal{B} \subseteq \mathcal{Z}$,

$$F(\mathcal{A}) + F(\mathcal{B}) \geq F(\mathcal{A} \cup \mathcal{B}) + F(\mathcal{A} \cap \mathcal{B}). \quad (4)$$

F is modular if strict equality holds in Equation 4, while F is monotone if for every $\mathcal{A} \subseteq \mathcal{B}$, $F(\mathcal{A}) \leq F(\mathcal{B})$.

B MONOTONICITY OF THE DIVERSIFICATION OBJECTIVE

Algorithm/Method	Monotonicity
[6, 34]	No
[2, 22, 23, 27, 35–37]	Yes

Table 2: Monotonicity of diversification algorithms.

C EXPERIMENTAL RESULTS

The plots of our experimental study is given in Figure 1. The plots show the trade-off between different relevance diversity metrics used in our experiments. Further details are given in section 5.

D BALANCED TRADE-OFF IN SPECIFIC DOMAINS AND ALGORITHMIC BIAS

As we said earlier, it is already established that the majority of the data is generated by a small set of users and follows Zipf's law, also known as the wisdom of a few [4]. Hence, there is an inherent

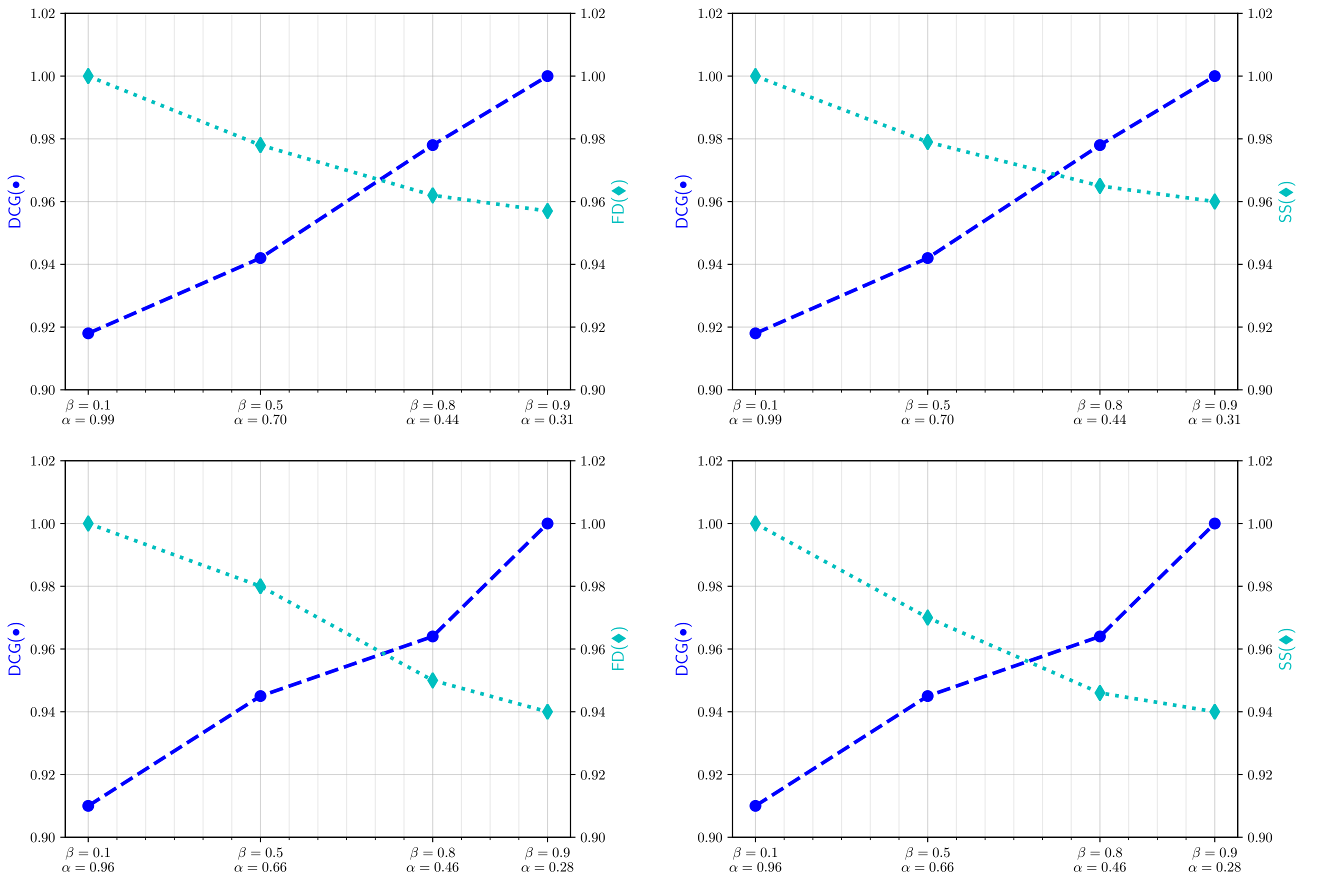

Figure 1: Relevance-Diversity trade-off for different values of α and β. Top and bottom rows represent the relevance-diversity trade-off for the algorithms in [27] and [2], respectively.

selection bias in the logged data and any learning algorithm trained on such data will inherit such a bias leading to algorithmic bias. Diversification approaches deal with algorithmic bias by countering popularity bias [31] and thus promoting rare items.

In sensitive domains like news recommendation, providing a balanced set of recommendations is vital. In typical news recommendation settings, users are expected to have inherent self-selection bias, as users mostly read news articles related to topics they are interested in. By providing the users with the most relevant recommendations only, with respect to a user-provided search query, the system is limiting users' *general view* about specific topics. For example, if a user reads articles which support extreme radical ideas, a standard recommendation strategy will keep on recommending more such articles and thus the reader will develop a very biased opinion about the topic. Hence, a standard recommendation strategy can exacerbate the self-selection bias already existing in domains like news recommendation.

An unbiased recommendation engine is expected to minimise polarisation in society, while informing and entertaining people, hence such a system should support diverse recommendations to make the people aware of orthogonal views. A relatively diverse recommendation will be the best bet against algorithmic bias, but it might lead to user dissatisfaction with consequential revenue loss. A well balanced recommendation will have a balanced mix of most relevant and diverse articles.

A balanced trade-off in relevance and diversity can also enhance the topical coverage a user is exposed to. For example, a user who is mostly interested in soccer will be recommended mostly soccer related news. By providing a balanced set of recommendations, users will receive more *off-topic* news, for example tennis or hockey, which might hinder the short term reward of the recommendations. But in the longer run, this will help the users to get a *broadened view* of the diverse topics.