Modified Group Generalized Binary Search with Near-Optimal Performance Guarantees

Gowtham Bellala, Clayton Scott
Department of Electrical Engineering and Computer Science
University of Michigan, Ann Arbor, MI 48109
{gowtham,clayscot} @umich.edu

December 3, 2010

Abstract

Group Generalized Binary Search (Group GBS) is an extension of the well known greedy algorithm GBS, for identifying the group of an unknown object while minimizing the number of binary questions posed about that object. This problem referred to as group identification or the Equivalence Class determination problem arises in applications such as disease diagnosis, toxic chemical identification, and active learning under persistent noise. Here, we propose a modified version of Group GBS and prove that it is competitive with the optimal algorithm. Our result holds even in the case where the queries have unequal costs.

1 Introduction

In applications such as active learning [1, 2, 3, 4], disease/fault diagnosis [5, 6, 7], toxic chemical identification [8], computer vision [9, 10] or the adaptive traveling salesman problem [11], one often encounters the problem of identifying an unknown object while minimizing the number of binary questions posed about that object. In these problems, there is a set $\Theta = \{\theta_1, \cdots, \theta_M\}$ of M different objects and a set $Q = \{q_1, \cdots, q_N\}$ of N distinct subsets of Θ known as queries. An unknown object θ is generated from this set Θ with a certain *prior* probability distribution $\Pi = (\pi_1, \cdots, \pi_M)$, i.e., $\pi_i = \Pr(\theta = \theta_i)$, and the goal is to uniquely identify this unknown object through as few queries from Q as possible, where a query $q \in Q$ returns a value 1 if $\theta \in q$, and 0 otherwise. For example, in active learning, the objects are classifiers and the queries are the labels for fixed test points. In active diagnosis, objects may correspond to faults, and queries to alarms. This problem has been generically referred to as binary testing or object/entity identification in the literature [5, 12].

The goal in object identification is to construct an optimal binary decision tree, where each internal node in the tree is associated with a query from Q, and each leaf node corresponds to an object from Θ. Optimality is often with respect to the expected depth of the leaf node corresponding to the unknown object θ. In general the determination of an optimal tree is NP-complete [13]. Hence, various greedy algorithms [5, 14] have been proposed to obtain a suboptimal binary decision tree. A well studied algorithm for this problem is known as the *splitting algorithm* [5] or *generalized binary search* (GBS) [1, 2]. This is the greedy algorithm which selects a query that most evenly divides the probability mass of the remaining objects [1, 2, 5, 15].

GBS assumes that the end goal is to rapidly identify individual objects. However, in applications such as disease diagnosis or toxic chemical identification, where Θ is a collection of possible diseases (or toxic chemicals), it may only be necessary to identify the intervention or response to an object, rather than the object itself. In these problems, the object set Θ is partitioned into groups and it is only necessary to identify the group to which the unknown object belongs. It has been noted that GBS is not necessarily efficient for group identification [16, 17].

In [16], we proposed an extension of GBS to the problem of group identification referred to as Group GBS and demonstrated the improved performance of Group GBS over GBS on synthetic datasets. Similarly, Golovin et.al. [17] studied the problem of group identification in the context of object identification under persistent noise, where they proposed an extension of an algorithm by Dasgupta [18] to the problem of group identification. In addition, they also show that their algorithm, referred to as Equivalence Class Edge Cutting (EC^2) constructs a tree whose cost(average depth) is logarithmically close to the optimal (least) cost when the prior probabilities are all rationals.

In this paper, we propose a modified version of Group GBS that achieves a logarithmic approximation to the optimal solution for any prior distribution. In addition, the upper bound achieved by the proposed algorithm is slightly better than that of EC^2.

1.1 Notation

We denote a group identification problem by $(\mathbf{B}, \Pi, \mathbf{y})$ where $\mathbf{B}$ is a known $M \times N$ binary matrix with b_{ij} equal to 1 if $\theta_i \in q_j$, and 0 otherwise. The vector $\mathbf{y} = (y_1, \cdots, y_M)$ denotes the group labels of the objects, where $y_i \in \{1, \cdots, K\}$, $K \leq M$. A decision tree T constructed on $(\mathbf{B}, \Pi, \mathbf{y})$ has a query from the set Q at each of its internal nodes, with the leaf nodes terminating in the objects from Θ. For a decision tree with L leaves, the leaf nodes are indexed by the set $\mathcal{L} = \{1, \cdots, L\}$ and the internal nodes are indexed by the set $\mathcal{I} = \{L+1, \cdots, 2L-1\}$. At any node '$a$', let $Q_a \subseteq Q$ denote the set of queries that have been performed along the path from the root node up to that node. Also, at any internal node $a \in \mathcal{I}$, let $l(a), r(a)$ denote the "left" and "right" child nodes, and let $\Theta_a \subseteq \Theta$ denote the set of objects that reach node 'a'. Thus, the sets $\Theta_{l(a)} \subseteq \Theta_a, \Theta_{r(a)} \subseteq \Theta_a$ correspond to the objects in Θ_a that respond 0 and 1 to the query at node 'a', respectively. We denote by $\pi_a := \sum_{\{i:\theta_i \in \Theta_a\}} \pi_i$, the probability mass of the objects reaching node 'a' in the tree.

In addition, let $\{\Theta^k\}_{k=1}^K$ be the partition of Θ, where $\Theta^k = \{\theta_i \in \Theta : y_i = k\}$. Then, Θ_a^k denotes the objects at node 'a' that belong to group k, and $\pi_a^k := \sum_{\{i:\theta_i \in \Theta_a^k\}} \pi_i$ denotes the probability mass of the objects in Θ_a^k. For a decision tree T constructed on $(\mathbf{B}, \Pi, \mathbf{y})$, and for any $\theta \in \Theta$, let $Q(T, \theta)$ denote the queries along the path from the root node to the leaf node ending in object θ. Then, the cost of identifying the group of an unknown object using the tree T is given by $c_{\text{avg}}(T) := \mathbb{E}_\theta[|Q(T,\theta)|] = \sum_{i=1}^M \pi_i |Q(T, \theta_i)|$.

2 Modified GGBS

Group GBS (GGBS) is a top-down, greedy algorithm that minimizes the expected number of queries required to identify the group of an unknown object θ [16]. At any internal node 'a' in the tree, the algorithm chooses a query that maximizes

$$H(\rho_a) - \sum_{k=1}^K \frac{\pi_a^k}{\pi_a} H(\rho_a^k) \tag{1}$$

where $\rho_a := \max\{\pi_{l(a)}, \pi_{r(a)}\}/\pi_a$, $\rho_a^k := \max\{\pi_{l(a)}^k, \pi_{r(a)}^k\}/\pi_a^k$ and $H(\rho) := -\rho \log_2 \rho - (1-\rho)\log_2(1-\rho)$ is the binary entropy function. Though GGBS performs significantly better than GBS on synthetic datasets as shown in [16], it is possible to construct datasets where it can perform significantly worse than the optimal solution [17].

Here, we present a modified version of GGBS and show that the proposed algorithm achieves a logarithmic approximation to the optimal solution. The new algorithm is to construct a top-down, greedy decision tree where at each internal node, a query that maximizes

$$\pi_{l(a)}\pi_{r(a)} - \sum_{k=1}^K \frac{\pi_a^k}{\pi_a} \pi_{l(a)}^k \pi_{r(a)}^k \tag{2}$$

is chosen, where the binary entropy terms $H(\rho_a)$ and $H(\rho_a^k)$ in (1) are approximated by the weighted Gini indices, $\pi_a^2(\rho_a(1-\rho_a))$ and $\left(\pi_a^k\right)^2(\rho_a^k(1-\rho_a^k))$, respectively. Note that in the special case where each group is of size 1, the query selection criterion in (2) reduces to $\pi_{l(a)}\pi_{r(a)}$, thereby reducing the proposed algorithm to GBS.

Given a group identification problem $(\mathbf{B}, \Pi, \mathbf{y})$, let $\mathcal{T}(\mathbf{B}, \Pi, \mathbf{y})$ denote the set of all possible trees that can uniquely identify the group of any object from the set Θ. Then, let T^* denote a tree with the least expected depth, i.e.,

$$T^* \in \underset{T \in \mathcal{T}(\mathbf{B}, \Pi, \mathbf{y})}{\arg\min}\ c_{\text{avg}}(T),$$

and let $\widehat{T}$ denote a tree constructed using the proposed greedy algorithm. The following theorem states that the expected depth of $\widehat{T}$ is logarithmically close to that of an optimal tree.

Theorem 1. *Let $(\mathbf{B}, \Pi, \mathbf{y})$ denote a group identification problem. For a greedy decision tree $\widehat{T}$ constructed using the greedy policy in (2), it holds that*

$$c_{avg}(\widehat{T}) \leq \left(2ln\left(\frac{1}{\sqrt{3}\pi_{min}}\right) + 1\right) c_{avg}(T^*) \tag{3}$$

where $\pi_{min} := \min_i \pi_i$ is the minimum prior probability of any object, and T^* is an optimal tree for the given problem.

Proof. In Appendix B. □

2.1 Unequal Query costs

Consider a group identification problem where the query costs are not the same for all queries in Q. Particularly, let $c(q)$ denote the cost incurred by querying q. Then, for a tree T constructed on a group identification problem $(\mathbf{B}, \Pi, \mathbf{y})$, the expected cost of identifying the group of an unknown object θ using the tree T is given by $c_{\text{avg}}(T) = \mathbb{E}_\theta[c(Q(T,\theta))]$, where $c(Q(T,\theta)) = \sum_{q \in Q(T,\theta)} c(q)$. Once again, an optimal tree corresponds to the tree with the least cost $c_{\text{avg}}(T)$.

The top-down, greedy decision tree is now constructed by choosing a query that maximizes $\frac{\Delta_a(q)}{c(q)}$ at each internal node, where $\Delta_a(q)$ is as defined in (2). Note that the quantity defined in (2) is a function of the query chosen, as $\pi_{l(a)}, \pi_{l(a)}^k, \pi_{r(a)}$ and $\pi_{r(a)}^k$ depends on the query chosen at node 'a' in the tree. The algorithm can be summarized as shown in Figure 1. Finally, the result in Theorem 1 holds for the case of unequal costs as well.

2.2 Simulation Results

We compare the performance of GBS, GGBS and modified GGBS on synthetic datasets constructed using the random data model of [16]. Note from Figure 2.1 that the modified GGBS algorithm performs significantly better than GBS and almost similar to that of GGBS.

3 Conclusions

We study the problem of group identification where the objects are partitioned into groups, and the goal is to identify the group of an unknown object using as few binary queries as possible. We present a modified version of a previously proposed algorithm known as Group GBS, and show that the modified algorithm achieves a logarithmic approximation to its optimal counterpart, under arbitrary prior distribution Π. We also extend the algorithm and the approximation result to the case of unequal query costs, where the goal is to minimize the expected cost of identifying the group of an unknown object.

Modified GGBS

Initialize: $\mathcal{L} = \{root\ node\}$, $Q_{root} = \emptyset$
while *some* $a \in \mathcal{L}$ *has more than one group*
  Choose query $q^* = \arg\max_{q \notin Q_a} \frac{\Delta_a(q)}{c(q)}$
  Form child nodes $l(a), r(a)$
  Replace 'a' with $l(a), r(a)$ in $\mathcal{L}$
end

$$\Delta_a(q) = \pi_{l(a)}\pi_{r(a)} - \sum_{k=1}^{K} \frac{\pi_a^k}{\pi_a} \pi_{l(a)}^k \pi_{r(a)}^k$$

Figure 1: Greedy algorithm for group identification with unequal query costs

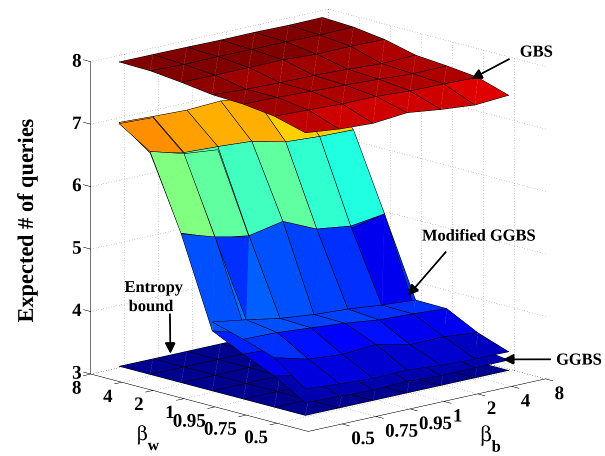

Figure 2: Expected number of queries required to identify the group of an unknown object using different algorithms.

Acknowledgements

G. Bellala and C. Scott were supported in part by NSF Awards No. 0830490 and 0953135.

References

[1] S. Dasgupta, "Analysis of a greedy active learning strategy," *Advances in Neural Information Processing Systems*, 2004.

[2] R. Nowak, "Generalized binary search," *Proceedings of the 46th Allerton Conference on Communications, Control and Computing*, pp. 568–574, 2008.

[3] ——, "Noisy generalized binary search," *Advances in Neural Information Processing Systems*, vol. 22, pp. 1366–1374, 2009.

[4] D. Golovin and A. Krause, "Adaptive Submodularity: A new approach to active learning and stochastic optimization," *In Proceedings of International Conference on Learning Theory (COLT)*, 2010.

[5] D. W. Loveland, "Performance bounds for binary testing with arbitrary weights," *Acta Informatica*, 1985.

[6] F. Yu, F. Tu, H. Tu, and K. Pattipati, "Multiple disease (fault) diagnosis with applications to the QMR-DT problem," *Proceedings of IEEE International Conference on Systems, Man and Cybernetics*, vol. 2, pp. 1187–1192, October 2003.

[7] J. Shiozaki, H. Matsuyama, E. O'Shima, and M. Iri, "An improved algorithm for diagnosis of system failures in the chemical process," *Computational Chemical Engineering*, vol. 9, no. 3, pp. 285–293, 1985.

[8] S. Bhavnani, A. Abraham, C. Demeniuk, M. Gebrekristos, A. Gong, S. Nainwal, G. Vallabha, and R. Richardson, "Network analysis of toxic chemicals and symptoms: Implications for designing first-responder systems," *Proceedings of American Medical Informatics Association*, 2007.

[9] D. Geman and B. Jedynak, "An active testing model for tracking roads in satellite images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 1, pp. 1–14, 1996.

[10] M. J. Swain and M. A. Stricker, "Promising directions in active vision," *International Journal of Computer Vision*, vol. 11, no. 2, pp. 109–126, 1993.

[11] A. Gupta, R. Krishnaswamy, V. Nagarajan, and R. Ravi, "Approximation algorithms for optimal decision trees and adaptive TSP problems," 2010, available online at arXiv.org:1003.0722.

[12] M. Garey, "Optimal binary identification procedures," *SIAM Journal on Applied Mathematics*, vol. 23(2), pp. 173–186, 1972.

[13] L. Hyafil and R. Rivest, "Constructing optimal binary decision trees is NP-complete," *Information Processing Letters*, vol. 5(1), pp. 15–17, 1976.

[14] S. R. Kosaraju, T. M. Przytycka, and R. S. Borgstrom, "On an optimal split tree problem," *Proceedings of 6th International Workshop on Algorithms and Data Structures, WADS*, pp. 11–14, 1999.

[15] R. M. Goodman and P. Smyth, "Decision tree design from a communication theory standpoint," *IEEE Transactions on Information Theory*, vol. 34, no. 5, 1988.

[16] G. Bellala, S. K. Bhavnani, and S. K. Bhavnani, "Extensions of Generalized Binary Search to Group Identification and Exponential costs," *Advances in Neural Information Processing Systems (NIPS)*, 2010.

[17] D. Golovin, D. Ray, and A. Krause, "Near-optimal Bayesian active learning with noisy observations," *Advances in Neural Information Processing Systems (NIPS)*, 2010.

[18] S. Dasgupta, "Coarse sample complexity bounds for active learning," *Advances in Neural Information Processing Systems*, 2006.

Appendix A. Short Review on Adaptive Submodularity

We briefly review the concept of adaptive submodularity introduced by Golovin et.al. [4]. Let $f : 2^Q \times \Theta \rightarrow \mathbb{R}_{\geq 0}$ be a utility/reward function that depends on the queries chosen and the unknown object $\theta \in \Theta$. Given a tree T, the expected cost of a tree is given by $c_{\text{avg}}(T) = \mathbb{E}_\theta[c(Q(T, \theta))]$. Given an $S > 0$, the optimal tree T^* is defined to be

$$T^* = \arg\min_T c_{\text{avg}}(T) \text{ such that } f(Q(T, \theta), \theta) \geq S, \ \forall \theta \in \Theta$$

Finding an optimal tree T^* is NP-complete and hence we need to resort to greedy approaches.

Definition 1. (Conditional Expected Marginal Gain) *Given the responses $\mathbf{z}_\mathcal{A}$ to previously chosen queries $Q_\mathcal{A}$, the* conditional expected marginal gain *of choosing a new query $q \notin Q_\mathcal{A}$ is given by*

$$\Delta(q|\mathbf{z}_\mathcal{A}) := \mathbb{E}_\theta[f(Q_\mathcal{A} \cup \{q\}, \theta) - f(Q_\mathcal{A}, \theta)|\mathbf{Z}_\mathcal{A} = \mathbf{z}_\mathcal{A}], \tag{4}$$

where the expectation is taken with respect to Π.

The greedy algorithm constructs a decision tree in a top-down manner, where at each stage a query that maximizes $\Delta(q|\mathbf{z}_\mathcal{A})/c(q)$ is chosen, i.e. $\arg\max_{q \notin Q_\mathcal{A}} \Delta(q|\mathbf{z}_\mathcal{A})/c(q)$.

Definition 2. (Strong Adaptive Monotonicity) *A function $f : 2^Q \times \Theta \rightarrow \mathbb{R}_{\geq 0}$ is* strongly adaptive monotone *with respect to* Π *if, informally "selecting more queries never hurts" with respect to the expected reward. Formally, for all $Q_\mathcal{A} \subseteq Q$, all $q \notin Q_\mathcal{A}$ and all $z \in \{0, 1\}$ such that* $\Pr(Z = z|\mathbf{Z}_\mathcal{A} = \mathbf{z}_\mathcal{A}) > 0$*, we require*

$$\mathbb{E}_\theta[f(Q_\mathcal{A}, \theta)|\mathbf{Z}_\mathcal{A} = \mathbf{z}_\mathcal{A}] \leq \mathbb{E}_\theta[f(Q_\mathcal{A} \cup \{q\}, \theta)|\mathbf{Z}_\mathcal{A} = \mathbf{z}_\mathcal{A}, Z = z]. \tag{5}$$

Definition 3. (Adaptive Submodular) *A function $f : 2^Q \times \Theta \rightarrow \mathbb{R}_{\geq 0}$ is adaptive submodular with respect to distribution Π if the conditional expected marginal gain of any fixed query does not increase as more queries are selected and their responses are observed. Formally, f is adaptive submodular w.r.t. Π if for all $Q_{\mathcal{A}}$ and $Q_{\mathcal{B}}$ such that $Q_{\mathcal{A}} \subseteq Q_{\mathcal{B}} \subseteq Q$ and for all $q \notin Q_{\mathcal{B}}$, we have*

$$\Delta(q|\mathbf{z}_{\mathcal{B}}) \leq \Delta(q|\mathbf{z}_{\mathcal{A}}) \tag{6}$$

Theorem 2. *Suppose $f : 2^Q \times \Theta \rightarrow \mathbb{R}_{\geq 0}$ is adaptive submodular and strongly adaptive monotone with respect to Π and there exists S such that $f(Q, \theta) = S$ for all $\theta \in \Theta$. Let η be any value such that $f(Q_{\mathcal{A}}, \theta) > S - \eta$ implies $f(Q_{\mathcal{A}}, \theta) = S$ for all $Q_{\mathcal{A}} \subseteq Q$ and all θ. Let T^* be an optimal tree with the least expected cost and let $\widehat{T}$ be a suboptimal tree constructed using the greedy algorithm, then*

$$c_{avg}(\widehat{T}) \leq c_{avg}(T^*) \left(ln \left(\frac{S}{\eta} \right) + 1 \right) \tag{7}$$

Appendix B. Proof of Theorem 1

Define the utility function to be $f(Q_{\mathcal{A}}, \theta_i) := 1 - \pi_a^2 + \left(\pi_a^{k_i}\right)^2$, where π_a is the probability mass of the objects remaining after observing responses to queries in $Q_{\mathcal{A}}$ with θ_i as the unknown object, and k_i denotes the group to which θ_i belongs. Note that $f(Q, \theta) = 1,\ \forall \theta \in \Theta$. Also, for any $Q_{\mathcal{A}} \subseteq Q$, if $f(Q_{\mathcal{A}}, \theta_i) > 1 - 3\pi_{\min}^2$, it implies $f(Q_{\mathcal{A}}, \theta_i) = 1$, hence $\eta = 3\pi_{\min}^2$. It also follows from Lemma 1 and Lemma 2 that f is adaptive submodular and strongly adaptive monotone. Hence, the result follows from Theorem 2 in Appendix A.

Lemma 1. *The objective function f defined above is adaptive submodular.*

Proof. Consider two subsets of Q such that $Q_{\mathcal{A}} \subseteq Q_{\mathcal{B}}$. Let $\mathbf{z}_{\mathcal{A}}, \mathbf{z}_{\mathcal{B}}$ denote the responses to the queries in $Q_{\mathcal{A}}$ and $Q_{\mathcal{B}}$, respectively. Then, we need to show that for any $q \notin Q_{\mathcal{B}}$, $\Delta(q|\mathbf{z}_{\mathcal{A}}) \geq \Delta(q|\mathbf{z}_{\mathcal{B}})$.

Let $\Theta_a \subseteq \Theta$ denote the set of objects whose responses to queries in $Q_{\mathcal{A}}$ are same as those in $\mathbf{z}_{\mathcal{A}}$. Then substituting $f(Q_{\mathcal{A}}, \theta) = 1 - \pi_a^2 + (\pi_a^k)^2$ in (4), we get

$$\begin{aligned}
\Delta(q|\mathbf{z}_{\mathcal{A}}) &= \sum_{k=1}^{K} \frac{\pi_{l(a)}^k}{\pi_a} \left[\pi_a^2 - \pi_{l(a)}^2 - (\pi_a^k)^2 + (\pi_{l(a)}^k)^2 \right] + \sum_{k=1}^{K} \frac{\pi_{r(a)}^k}{\pi_a} \left[\pi_a^2 - \pi_{r(a)}^2 - (\pi_a^k)^2 + (\pi_{r(a)}^k)^2 \right] \\
&= \frac{\pi_{l(a)}}{\pi_a} \pi_{r(a)} (\pi_a + \pi_{l(a)}) - \sum_{k=1}^{K} \frac{\pi_{l(a)}^k}{\pi_a} \pi_{r(a)}^k (\pi_a^k + \pi_{l(a)}^k) + \frac{\pi_{r(a)}}{\pi_a} \pi_{l(a)} (\pi_a + \pi_{r(a)}) - \sum_{k=1}^{K} \frac{\pi_{r(a)}^k}{\pi_a} \pi_{l(a)}^k (\pi_a^k + \pi_{r(a)}^k) \\
&= 3\pi_{l(a)} \pi_{r(a)} - \sum_{k=1}^{K} 3 \frac{\pi_a^k}{\pi_a} \pi_{l(a)}^k \pi_{r(a)}^k.
\end{aligned}$$

Similarly, let $\Theta_b \subseteq \Theta$ denote the set of objects whose responses to queries in $Q_{\mathcal{B}}$ are equal to those in $\mathbf{z}_{\mathcal{B}}$. Then, substituting $f(Q_{\mathcal{B}}, \theta) = 1 - \pi_b^2 + (\pi_b^k)^2$ in (4), we get $\Delta(q|\mathbf{z}_{\mathcal{B}}) = 3\pi_{l(b)} \pi_{r(b)} - \sum_{k=1}^{K} 3 \frac{\pi_b^k}{\pi_b} \pi_{l(b)}^k \pi_{r(b)}^k$.

To prove f is adaptive submodular, we need to show that

$$\begin{aligned}
\pi_{l(a)} \pi_{r(a)} - \sum_{k=1}^{K} \frac{\pi_a^k}{\pi_a} \pi_{l(a)}^k \pi_{r(a)}^k &\geq \pi_{l(b)} \pi_{r(b)} - \sum_{k=1}^{K} \frac{\pi_b^k}{\pi_b} \pi_{l(b)}^k \pi_{r(b)}^k, \\
\implies \pi_a \pi_b \pi_{l(a)} \pi_{r(a)} - \sum_{k=1}^{K} \pi_a^k \pi_b \pi_{l(a)}^k \pi_{r(a)}^k &\geq \pi_a \pi_b \pi_{l(b)} \pi_{r(b)} - \sum_{k=1}^{K} \pi_b^k \pi_a \pi_{l(b)}^k \pi_{r(b)}^k
\end{aligned}$$

Note that since $Q_{\mathcal{A}} \subseteq Q_{\mathcal{B}}$, $\Theta_b \subseteq \Theta_a$ and hence $\pi_b \leq \pi_a$, $\pi_b^k \leq \pi_a^k$, $\forall k$. For any query $q \notin Q_{\mathcal{B}}$, let $\Theta_{l(a)}$ and $\Theta_{r(a)}$ correspond to the objects in Θ_a that respond 0 and 1 to query q respectively. Similarly, let $\Theta_{l(b)}$ and

$\Theta_{r(b)}$ correspond to the objects in Θ_b that respond 0 and 1 to query q respectively. Then, $\pi_{l(b)} \leq \pi_{l(a)}$, $\pi^k_{l(b)} \leq \pi^k_{l(a)}$, $\forall k$, and $\pi_{r(b)} \leq \pi_{r(a)}$, $\pi^k_{r(b)} \leq \pi^k_{r(a)}$, $\forall k$. Hence

$$\pi_a\pi_b\pi_{l(a)}\pi_{r(a)} - \sum_{k=1}^{K} \pi^k_a\pi_b\pi^k_{l(a)}\pi^k_{r(a)} = \pi_a\pi_b \sum_{k=1}^{K} \pi^k_{l(a)}\pi^k_{r(a)} + \pi_a\pi_b \sum_{k\neq m} \pi^k_{l(a)}\pi^m_{r(a)} - \sum_{k=1}^{K} \pi^k_a\pi_b\pi^k_{l(a)}\pi^k_{r(a)}$$

$$= \sum_{k=1}^{K} \pi^k_{l(a)}\pi^k_{r(a)}(\pi_a - \pi^k_a)\pi_b + \pi_a\pi_b \sum_{k\neq m} \pi^k_{l(a)}\pi^m_{r(a)} \tag{8a}$$

$$\geq \sum_{k=1}^{K} \pi^k_{l(a)}\pi^k_{r(a)}(\pi_a - \pi^k_a)\pi_b + \pi_a\pi_b \sum_{k\neq m} \pi^k_{l(b)}\pi^m_{r(b)} \tag{8b}$$

$$= \sum_{k=1}^{K} \pi^k_{l(a)}\pi^k_{r(a)}(\pi_a - \pi^k_a)(\pi_b - \pi^k_b) + \sum_{k=1}^{K} \pi^k_{l(a)}\pi^k_{r(a)}(\pi_a - \pi^k_a)\pi^k_b + \pi_a\pi_b \sum_{k\neq m} \pi^k_{l(b)}\pi^m_{r(b)} \tag{8c}$$

$$\geq \sum_{k=1}^{K} \pi^k_{l(b)}\pi^k_{r(b)}(\pi_a - \pi^k_a)(\pi_b - \pi^k_b) + \sum_{k=1}^{K} \pi^k_{l(a)}\pi^k_{r(a)}(\pi_a - \pi^k_a)\pi^k_b + \pi_a\pi_b \sum_{k\neq m} \pi^k_{l(b)}\pi^m_{r(b)} \tag{8d}$$

$$\geq \sum_{k=1}^{K} \pi^k_{l(b)}\pi^k_{r(b)}(\pi_a - \pi^k_a)(\pi_b - \pi^k_b) + \sum_{k=1}^{K} \pi^k_{l(b)}\pi^k_{r(b)}(\pi_b - \pi^k_b)\pi^k_a + \pi_a\pi_b \sum_{k\neq m} \pi^k_{l(b)}\pi^m_{r(b)} \tag{8e}$$

$$= \sum_{k=1}^{K} \pi^k_{l(b)}\pi^k_{r(b)}\pi_a(\pi_b - \pi^k_b) + \pi_a\pi_b \sum_{k\neq m} \pi^k_{l(b)}\pi^m_{r(b)}$$

$$= \pi_a\pi_b\pi_{l(b)}\pi_{r(b)} - \sum_{k=1}^{K} \pi_a\pi^k_b\pi^k_{l(b)}\pi^k_{r(b)}$$

where (8e) follows from (8d) since

$$\begin{aligned}\sum_{k=1}^{K} \pi^k_{l(a)}\pi^k_{r(a)}(\pi_a - \pi^k_a)\pi^k_b &= \sum_{k=1}^{K} \pi^k_{l(a)}\pi^k_{l(b)}\pi^k_{r(a)}(\pi_a - \pi^k_a) + \sum_{k=1}^{K} \pi^k_{r(a)}\pi^k_{r(b)}\pi^k_{l(a)}(\pi_a - \pi^k_a)\\ &\geq \sum_{k=1}^{K} \pi^k_{l(a)}\pi^k_{l(b)}\pi^k_{r(b)}(\pi_b - \pi^k_b) + \sum_{k=1}^{K} \pi^k_{r(a)}\pi^k_{r(b)}\pi^k_{l(b)}(\pi_b - \pi^k_b)\\ &= \sum_{k=1}^{K} \pi^k_{l(b)}\pi^k_{r(b)}(\pi_b - \pi^k_b)\pi^k_a,\end{aligned}$$

thus proving that f is adaptive submodular. □

Lemma 2. *The objective function f is strongly adaptive monotone.*

Proof. Consider any subset of queries $Q_{\mathcal{A}} \subseteq Q$, and let $\mathbf{z}_{\mathcal{A}}$ denote the responses to these queries. Let Θ_a denote the set of objects whose responses to queries in $Q_{\mathcal{A}}$ are equal to those of $\mathbf{z}_{\mathcal{A}}$. For any query $q \notin Q_{\mathcal{A}}$, let $\Theta_{l(a)}$ and $\Theta_{r(a)}$ correspond to the objects in Θ_a that respond 0 and 1 to query q respectively.

For strong adaptive monotonicity, we need to show that

$$1 - \pi_a^2 + \sum_{k=1}^{K} \frac{(\pi^k_a)^3}{\pi_a} \leq 1 - \pi^2_{l(a)} + \sum_{k=1}^{K} \frac{(\pi^k_{l(a)})^3}{\pi_{l(a)}}, \quad \text{if } \pi_{l(a)} > 0$$

$$\text{and } 1 - \pi_a^2 + \sum_{k=1}^{K} \frac{(\pi^k_a)^3}{\pi_a} \leq 1 - \pi^2_{r(a)} + \sum_{k=1}^{K} \frac{(\pi^k_{r(a)})^3}{\pi_{r(a)}}, \quad \text{if } \pi_{r(a)} > 0.$$

We will show the first inequality, and the second inequality can be shown in a similar manner. Given $\pi_{l(a)} > 0$, we need to show that

$$\pi_a^3 \pi_{l(a)} - \pi_{l(a)}^3 \pi_a \geq \sum_{k=1}^{K} (\pi_a^k)^3 \pi_{l(a)} - (\pi_{l(a)}^k)^3 \pi_a.$$

Note that

$$\begin{aligned}
&\pi_a^3 \pi_{l(a)} - \pi_{l(a)}^3 \pi_a = (\pi_{l(a)} + \pi_{r(a)})^3 \pi_{l(a)} - \pi_{l(a)}^3 (\pi_{l(a)} + \pi_{r(a)}) \\
&= \pi_{r(a)}^3 \pi_{l(a)} + 3\pi_{l(a)}^2 \pi_{r(a)}^2 + 2\pi_{l(a)}^3 \pi_{r(a)} && \text{(9a)} \\
&\geq \sum_{k=1}^{K} \left[(\pi_{r(a)}^k)^3 \pi_{l(a)} + 3\pi_{l(a)} \pi_{l(a)}^k (\pi_{r(a)}^k)^2 \right] + 2\pi_{l(a)}^3 \pi_{r(a)} && \text{(9b)} \\
&= \sum_{k=1}^{K} \left[(\pi_{r(a)}^k)^3 \pi_{l(a)} + 3\pi_{l(a)} \pi_{l(a)}^k (\pi_{r(a)}^k)^2 - (\pi_{l(a)}^k)^3 \pi_{r(a)} \right] + 2\pi_{l(a)}^3 \pi_{r(a)} + \sum_{k=1}^{K} (\pi_{l(a)}^k)^3 \pi_{r(a)} && \text{(9c)} \\
&\geq \sum_{k=1}^{K} \left[(\pi_{r(a)}^k)^3 \pi_{l(a)} + 3\pi_{l(a)} \pi_{l(a)}^k (\pi_{r(a)}^k)^2 - (\pi_{l(a)}^k)^3 \pi_{r(a)} + 3(\pi_{l(a)}^k)^2 \pi_{r(a)}^k \pi_{l(a)} \right] && \text{(9d)} \\
&= \sum_{k=1}^{K} \left[(\pi_{l(a)}^k)^3 + 3(\pi_{l(a)}^k)^2 \pi_{r(a)}^k + 3\pi_{l(a)}^k (\pi_{r(a)}^k)^2 + (\pi_{r(a)}^k)^3 \right] \pi_{l(a)} - (\pi_{l(a)}^k)^3 \pi_{l(a)} - (\pi_{l(a)}^k)^3 \pi_{r(a)} && \text{(9e)} \\
&= \sum_{k=1}^{K} (\pi_a^k)^3 \pi_{l(a)} - (\pi_{l(a)}^k)^3 \pi_a
\end{aligned}$$

where (9b) follows from (9a) as $\pi_{r(a)}^3 \pi_{l(a)}$, $3\pi_{l(a)} \pi_{l(a)} \pi_{r(a)}^2$ has more non-negative terms than $\sum_{k=1}^{K} (\pi_{r(a)}^k)^3 \pi_{l(a)}$, $\sum_{k=1}^{K} 3\pi_{l(a)} \pi_{l(a)}^k (\pi_{r(a)}^k)^2$, respectively. Also (9d) follows from (9c) since

$$\begin{aligned}
&\pi_{r(a)} \left[2\pi_{l(a)}^3 + \sum_{k=1}^{K} (\pi_{l(a)}^k)^3 \right] = \pi_{r(a)} \left[\sum_{k=1}^{K} 3(\pi_{l(a)}^k)^3 + 6 \sum_{k \neq m} (\pi_{l(a)}^k)^2 \pi_{l(a)}^m + 6 \sum_{k \neq m \neq n} \pi_{l(a)}^k \pi_{l(a)}^m \pi_{l(a)}^n \right] \\
&= \left(\sum_{j=1}^{K} \pi_{r(a)}^j \right) \left[\sum_{k=1}^{K} 3(\pi_{l(a)}^k)^3 + 6 \sum_{k \neq m} (\pi_{l(a)}^k)^2 \pi_{l(a)}^m + 6 \sum_{k \neq m \neq n} \pi_{l(a)}^k \pi_{l(a)}^m \pi_{l(a)}^n \right] \\
&\geq 3 \sum_{k=1}^{K} (\pi_{l(a)}^k)^3 \pi_{r(a)}^k + 3 \sum_{k \neq m} (\pi_{l(a)}^k)^2 \pi_{r(a)}^k \pi_{l(a)}^m \\
&= 3\pi_{l(a)} \sum_{k=1}^{K} (\pi_{l(a)}^k)^2 \pi_{r(a)}^k,
\end{aligned}$$

thus proving that f is strongly adaptive monotone. □