Appendices to "Bi-Criteria Diverse Plan Selection via Beam Search Approximation", published at SoCS 2024

Shanhe Zhong [1], Pouya Shati [2,3], Eldan Cohen[1]

[1] Department of Mechanical and Industrial Engineering, University of Toronto
[2] Department of Computer Science, University of Toronto
[3] Vector Institute, Toronto, Canada
{jszhong, ecohen}@mie.utoronto.ca, pouya@cs.utonronto.ca

A Proofs

We present the proofs that we omitted in the main paper. We also introduce two lemmas to help with the proofs. We provide the proof for Lemma 1 and refer the reader to [1] for the proof of Lemma 2.

Lemma 1. *The top solution in each iteration of the beam search procedure, has an equal or higher objective value compared to the best expansion of the top solution from last iteration.*

$$f(B_t^1) \geq f(B_{t-1}^1 \cup \{\pi\}), \forall \pi \in \Pi \setminus B_{t-1}.$$

Proof. By the mathematical definition of the update of beam search, we know $f(B_t^1) \geq f(B), \forall B \in B'_{t-1}$ as $f(B_t^1) = \max_{B \in B_t} f(B) = \max_{B \in B'_{t-1}} f(B)$. Since $B_{t-1}^1 \cup \{\pi\} \in B'_{t-1}$ for any $\pi \in \Pi \setminus B_{t-1}^1$, the lemma follows.

Lemma 2 (Ravi et al. [1]). *Given a metric distance function $d(\cdot,\cdot)$ and a set metric distance function $h(\cdot,\cdot)$ where $h(X,Y) = \sum_{x \in X, y \in Y} d(x,y)$ and $h(X,X) = 2h(X)$, we have the following inequality for disjoint sets X and Y:*

$$(|X| - 1)h(X,Y) \geq |Y|h(X).$$

Proposition 1. *Beam search that operates on the objective function $f(S) = g_{sum}(S)$ finds the optimal solution in both $g_{sum}(S) = \sum_{\pi_i \in S} c(\pi_i)$ and $g_{min}(S) = \min_{\pi_i \in S} c(\pi_i)$.*

Proof. It is important to realize that we are operating with scaled plan costs $c(\pi)$ as defined previously:

$$c(\pi) = 1 - \frac{C(\pi) - \min_{\pi \in \Pi}(C(\pi))}{\max_{\pi \in \Pi}(C(\pi)) - \min_{\pi \in \Pi}(C(\pi))}$$

In this way, the rank is preserved but reversed. Suppose a set $S^* = \operatorname{argmax}_{\pi_1,\ldots,\pi_k \in \Pi} \sum_{i=1}^k c(\pi_i)$ optimizes g_{sum}. Now, we show that S^* also optimizes g_{min}.

BWOC, assume there exists $S' \subset \Pi$ with $|S'| = k$ such that $g_{min}(S') > g_{min}(S^*)$. Then, since S^* and S' are of the same size, there exists $\pi' \in S' \setminus S^*$ such that $c(\pi^*) < c(\pi')$ where $\pi^* = \underset{\pi \in S^*}{\text{argmin}}\ c(\pi)$.

However, $g_{sum}(S^* \cup \{\pi'\} \setminus \{\pi^*\}) > g_{sum}(S^*)$ as $c(\pi') > c(\pi^*)$ ($\Rightarrow\Leftarrow$). This means there is no such S' and hence S^* necessarily optimizes g_{min} as well.

Therefore, as a set that maximizes g_{sum} also maximizes g_{min}, it suffices to show that beam search maximizes g_{sum}. Now, suppose $S_i^* = \underset{\pi_1,\ldots,\pi_k \in \Pi}{\text{argmax}} \sum_{j=1}^{i} c(\pi_j)$.

Proposition: The top solution at the end of iteration $i \in \{2, 3, ..., k\}$ is equivalent to S_i^*:

$$g_{sum}(B_i^1) = g_{sum}(S_i^*).$$

Base Case: This is trivially true as in our implementation of beam search, a pair with optimal cost is selected at first.

Inductive Hypothesis: Suppose the proposition holds for some $i-1$. That is, $g_{sum}(B_{i-1}^1) = g_{sum}(S_{i-1}^*)$. Now to prove by contradiction, assume that $g_{sum}(B_i^1) < g_{sum}(S_i^*)$. Consider $\pi' = \underset{\pi \in \Pi \setminus B_{i-1}^1}{\text{argmax}}\ c(\pi)$ and $B_i' = B_{i-1}^1 \cup \{\pi'\}$.

Note that $g_{sum}(B_i') \leq g_{sum}(B_i) < g_{sum}(S_i^*)$ and consider $\pi^{max} = \underset{\pi \in S_i^* \setminus B_{i-1}^1}{\text{argmax}}\ c(\pi)$.

Note that by definition $c(\pi') \geq c(\pi^{max})$. Thus, we have:

$$\begin{aligned} g_{sum}(B_i') &= g_{sum}(B_{i-1}^1) + c(\pi') \\ &< g_{sum}(S_i^* - \{\pi^{max}\}) + c(\pi^{max}) \end{aligned}$$

Therefore, we have $g_{sum}(B_{i-1}^1) < g_{sum}(S_i^* - \{\pi^{max}\}])$ which contracts B_{i-1}^1 being equivalent to the optimal solution.

Theorem 1. *For a diverse plan selection task of size k from the set Π, beam search obtains a $1/2$-approximation to objective function $f(S) = h_{min}(S)$, where $h_{min}(S) = \min_{\pi_i, \pi_j \in S} d(\pi_i, \pi_j)$ and d is a metric distance function.*

Proof. We augment the proof by [2].

Suppose $B_t = \{B_t^1, ..., B_t^m\}$ represents the beam at iteration t for $t \in \{2, ..., k\}$, with each partial solution in descending order by their h_{min}. That is, $h_{min}(B_t^i) \geq h_{min}(B_t^j), \forall 1 \leq i \leq j \leq m$.

Consider a subset with optimal minimum pairwise diversity $S^* = \{\pi_1^*, \pi_2^*, ..., \pi_k^*\}$ and $h_{min}(S^*) = h^*$. By definition, $\forall \pi_i^*, \pi_j^* \in S^*$, we have $d(\pi_i^*, \pi_j^*) \geq h^*$. Then, as $d(\cdot, \cdot)$ is a metric distance function, we can construct k open balls of radius $\frac{h^*}{2}$, and each open ball b_j is centered at π_j^*. By construction for each i, $b_i \cap S^* = \{\pi_i^*\}$ and for each pair $i \neq j, b_i \cap b_j = \emptyset$. Now, note that for a plan π, $\pi \notin b_j$ is equivalent to $d(\pi, \pi_j^*) \geq \frac{h^*}{2}$.

Since each pair of the k open balls has empty intersections and B_i^1 is of size i, there will be at least $k - i$ open balls left empty denoted by b, where $\forall b_j \in b$ and $\pi \in B_i^1$, $\pi \notin b_j$ which implies $d(\pi, \pi_j^*) \geq \frac{h^*}{2}$.

Proposition: The top solution of beam search at the end of iteration $i \in \{2, 3, ..., k\}$ satisfies: $h_{min}(B_i^1) \geq \frac{h^*}{2}$.

Base Case: This is trivially true as in our implementation of beam search, a pair with optimal pairwise diversity is selected at first.

Inductive Step: Suppose $h_{min}(B_i^1) \geq \frac{h^*}{2}$; need to show $h_{min}(B_{i+1}^1) \geq \frac{h^*}{2}$.

We previously showed that there are still at least $k - i$ plans π_j^* (center of an open ball) to choose from that satisfies $\forall \pi \in B_i^1$, $d(\pi, \pi_j^*) \geq \frac{h^*}{2}$. Say we pick any of the π_j^*. This implies $h_{min}(B_i^1 \cup \{\pi_j^*\}) \geq \frac{h^*}{2}$ still holds. Now, it follows from Lemma 1 that since $B_i^1 \cup \{\pi_j^*\}$ is a successor of B_i^1, we have $h_{min}(B_{i+1}^1) \geq h_{min}(B_i^1 \cup \{\pi_j^*\}) \geq \frac{h^*}{2}$. Therefore, the proposition follows from induction.

Theorem 2. *For a diverse plan selection task of size k from the set Π, beam search obtains a $1/2$-approximation to the function $f(S) = \alpha g(S) + \beta h_{sum}(S)$ for all monotone, submodular, and non-negative $g(S)$ and $\alpha, \beta \in \mathbb{R}$ where $h_{sum}(S) = \sum_{\pi_i, \pi_j \in S} d(\pi_i, \pi_j)$ and d is a metric distance function.*

Proof. We augment the proof by [3].

We prove the theorem for a beam search procedure that starts from partial solutions of size 0 rather than 2 as in our implementation. Note that we select optimal partial solutions of size 2, therefore the proof remains valid. Furthermore, we refer to h_{sum} as simply h due to brevity.

Consider S^* to be the optimal solution in f where $|S^*| = k$, S is the solution from beam search, and B_i^1 is the best subset of cardinality i selected by beam search at i-th iteration for some $i \leq k$. In particular, we have $S = B_k^1$.

For simplicity, we define $g'(\cdot) = \frac{1}{2}g(\cdot)$ and $f'(\cdot) = \alpha g'(\cdot) + \beta h(\cdot)$. Now, we would like to show the following inequality for any $i < k$:

$$f'(B_{i+1}^1) - f'(B_i^1) \geq \frac{1}{k}[\alpha g'(S^*) - \alpha g'(S)] + \frac{i\beta}{k(k-1)} h(S^*).$$

Let $A = S^* \cap B_i^1$, $B = B_i^1 - A$ and $C = S^* - A$. We prove the case $|C| = 1$ separately as it is a special case that can break the general proof for $|C| > 1$. Note that the proof for $|C| = 0$ is trivial as it indicates the output being the exact same as the optimal solution.

If $|C| = 1$, it means we have $i = k - 1$ and $B_i^1 \subset S^*$. Let c be the element in C and π be the best element taken by the beam search for its top candidate of iteration i. Let S be the result set at the end of the beam search. Then, by definition, the top successor will have a greater or equal gain, which is $f(B_i^1 \cup \{\pi\}) \geq f(B_i^1 \cup \{c\})$. Therefore, we have:

$$
\begin{aligned}
f(B_{i+1}^1) &\geq f'(B_{i+1}^1) \\
&\geq f'(B_i^1 \cup \{\pi\}) && \text{(by Lemma 1)} \\
&\geq f'(B_i^1 \cup \{c\}) \\
&= f'(S^*) \\
&\geq \frac{1}{2} f(S^*).
\end{aligned}
$$

Now, for the case of $|C| > 1$, we introduce some inequalities using Lemma 2:

$$(|C| - 1)h(B, C) \geq |B|h(C). \tag{1}$$

$$(|C| - 1)h(A, C) \geq |A|h(C). \tag{2}$$

$$(|A| - 1)h(A, C) \geq |C|h(A). \tag{3}$$

By the properties of set operations, we also have

$$h(A) + h(A, C) + h(C) = h(S^*). \tag{4}$$

We compute the linear combination of $\frac{1}{|C|-1} * (1) + \frac{|C|-|B|}{k(|C|-1)} * (2) + \frac{i}{k(k-1)} * (3) + \frac{i|C|}{k(k-1)} * (4)$ to get: $h(A, C) + h(B, C) - \frac{i|C|(k-|C|)}{k(k-1)(|C|-1)} h(C) \geq \frac{i|C|}{k(k-1)} h(S^*)$. Since $k > |C| \iff -\frac{i|C|(k-|C|)}{k(k-1)(|C|-1)} h(C) \leq 0$ and $h(A, C) + h(B, C) = h(B_i^1, C)$, we also have:

$$h(B_i^1, C) \geq \frac{i|C|}{k(k-1)} h(S^*). \tag{5}$$

By submodularity and monotonicity of g' (as submodularity is closed under linear combination and monotonicity is retained with a positive coefficient), we have:

$$
\begin{aligned}
&\sum_{x \in C} [g'(B_i^1 \cup \{x\}) - g'(B_i^1)] \\
&\geq g'(C \cup B_i^1) - g'(B_i^1) \\
&\geq g'(S^*) - g'(B_i^1).
\end{aligned} \tag{6}
$$

Therefore, we can compute:

$$\begin{aligned}
&\sum_{c\in C}[f'(B_i^1\cup\{c\})-f'(B_i^1)]\\
=&\sum_{c\in C}[\alpha g'(B_i^1\cup\{c\})-\alpha g'(B_i^1)+\beta h(\{c\},B_i^1)]\\
=&\sum_{c\in C}[\alpha g'(B_i^1\cup\{c\})-\alpha g'(B_i^1)]+\beta h(C,B_i^1)\\
\geq&[\alpha g'(S^*)-\alpha g'(B_i^1)]+\beta h(C,B_i^1) && \text{(by (6))}\\
\geq&[\alpha g'(S^*)-\alpha g'(B_i^1)]+\frac{i\beta|C|}{k(k-1)}h(S^*). && \text{(by (5))}
\end{aligned}$$

By Lemma 1, the marginal gain of $f'(B_{i+1}^1)$ over $f'(B_i^1)$ must be greater than or equal to B_i^1 in union with any of the choices in C; this implies it must also be greater than or equal to their average:

$$\begin{aligned}
&f'(B_{i+1}^1)-f'(B_i^1)\\
\leq&\frac{1}{|C|}[\alpha g'(S^*)-\alpha g'(B_i^1)]+\frac{i\beta}{k(k-1)}h(S^*)\\
\leq&\frac{1}{k}[\alpha g'(S^*)-\alpha g'(B_i^1)]+\frac{i\beta}{k(k-1)}h(S^*)\\
\leq&\frac{1}{k}[\alpha g'(S^*)-\alpha g'(S)]+\frac{i\beta}{k(k-1)}h(S^*) && \text{(by monotonicity)}
\end{aligned}$$

We proceed to sum up all the marginal gains from 0 to $k-1$. We see intermediate terms cancels out and are left with $f'(B_k^1)-f'(B_0^1)=f'(S)-f'(\emptyset)=f'(S)$. Thus:

$$\begin{aligned}
f'(S)&=\sum_{i=0}^{k-1}[f'(B_{i+1}^1)-f'(B_i^1)]\\
&\geq\alpha g'(S^*)-\alpha g'(S)+\frac{\beta}{2}h(S^*).
\end{aligned}$$

Therefore, this implies:

$$\begin{aligned}
\alpha g'(S)+\beta h(S)&\geq[\alpha g'(S^*)-\alpha g'(S)+\frac{\beta}{2}h(S^*)]\\
\alpha g(S)+\beta h(S)&\geq[\frac{\alpha}{2}g(S^*)+\frac{\beta}{2}h(S^*)]
\end{aligned}$$

Then, by definition, we rearrange to get:

$$\begin{aligned} f(S) &= \alpha g(S) + \beta h(S) \\ &\geq [\frac{\alpha}{2} g(S^*) + \frac{\beta}{2} h(S^*)] \\ &= \frac{1}{2} f(S^*). \end{aligned}$$

Corollary 1. *The best solution from two beam search runs, one optimizing $g(s)$ and the other optimizing $h(s)$, both guaranteeing an approximation ratio of at least $1/2$, is guaranteed to obtain a $1/4$-approximation to the objective function $f(S) = \alpha g(S) + \beta h(S)$ with $\alpha, \beta \in \mathbb{R}$.*

Proof. We augment the proof by [2].

To prove the approximation ratio for $f(S) = \alpha g(S) + \beta h(S)$, we run beam search twice. First, we run beam search with $\alpha = 0$ to optimize for diversity and denote the beam at the last iteration as B_h. Then, we run beam search for g and $\beta = 0$ to optimize for quality and denote the beam at the last iteration as B_g. Suppose O_h and O_g achieves optimal h and g respectively. It follows from the assumption that $f(B_h^1) \geq \beta h(B_h^1) \geq \beta h(O_h)/2$ and $f(B_g^1) \geq \alpha g(B_g^1) \geq \alpha g(O_g)/2$.

We set the beam search solution from the two runs to be X where $X = \underset{S \in B_g \cup B_h}{\operatorname{argmax}} f(S)$. Then, as $f(X) \geq f(B_g^1)$ and $f(X) \geq f(B_h^1)$ which implies $2f(X) \geq f(B_g^1) + f(B_h^1)$, it follows that:

$$\begin{aligned} f(X) &\geq \frac{(f(B_g^1) + f(B_h^1))}{2} \\ &\geq \frac{(\frac{\alpha g(O_g)}{2} + \frac{\beta h(O_h)}{2})}{2} \\ &\geq \frac{(\alpha g(O_g) + \beta h(O_h))}{4}. \end{aligned} \tag{1}$$

Suppose $O = \underset{S \subset \Pi}{\operatorname{argmax}} f(S)$ with $|O| = k$. By definition of O_h and O_g, it follows that $h(O) \leq h(O_h)$ and $g(O) \leq g(O_g)$. Then, after substituting this into (1), we see that $f(X) \geq (\alpha g(O) + \beta h(O))/4$, which completes our proof.

B Examples of Beam Search Underperforming Compared Greedy Search

We now present two simple cases where greedy search outperforms beam search. For simplicity, we show this for the beam search with a beam width of 2. We first define the candidate set to choose from: $\Pi = \{\pi_1, \pi_2, \pi_3, \pi_4, \pi_5, \pi_6, \pi_7\}$, where the set representations of the plans are: $\pi_1 = \{1, 1, 1, 1, 0\}$, $\pi_2 = \{1, 0, 1, 0, 1\}$, $\pi_3 = \{1, 0, 0, 0, 0\}$, $\pi_4 = \{1, 0, 1, 0, 0\}$, $\pi_5 = \{1, 1, 0, 0, 1\}$, $\pi_6 = \{1, 1, 0, 1, 1\}$, $\pi_7 = \{1, 1, 1, 0, 1\}$. For the following examples, we choose $k = 5$. In Example 1 we first show beam search underperforms on h_{min} compared to greedy search.

Example 1. At the first iteration, both algorithms start with one optimal pair, which is: $\{\pi_4, \pi_6\}$ with $h_{min}(\{\pi_4, \pi_6\}) = 0.8$ and beam search starts with an addition pair $\{\pi_2, \pi_3\}$ with $h_{min}(\{\pi_2, \pi_3\}) = 0.67$. In the next iteration, the successors of $\{\pi_2, \pi_3\}$ outperforms the successors of $\{\pi_4, \pi_6\}$ as $B_3 = [\{\pi_2, \pi_3, \pi_6\}, \{\pi_1, \pi_2, \pi_3\}]$ with $h_{min}(B_3^1)$ $= h_{min}(B_3^2) = 0.6$, while $G_3 = \{\pi_3, \pi_4, \pi_6\}$ has $h_{min}(G_3) = 0.5$. Though beam search gets a higher diversity at this iteration, the inclusion of π_2 results in an inferior selection. This is evident as $G_4 = \{\pi_1, \pi_3, \pi_4, \pi_6\}$ and $G_5 = \{\pi_1, \pi_3, \pi_4, \pi_6, \pi_7\}$ as $h_{min}(G_4) = h_{min}(G_5) = 0.5$, while for beam search, the diversity of solution $B_4 = [\{\pi_1, \pi_2, \pi_3, \pi_5\}, \{\pi_1, \pi_2, \pi_3, \pi_6\}]$ degrades as $h_{min}(B_4^1) = 0.5$ and $h_{min}(B_4^2) = 0.4$. In the end, $B_5 = [\{\pi_1, \pi_2, \pi_3, \pi_4, \pi_5\}, \{\pi_1, \pi_2, \pi_3, \pi_4, \pi_6\}]$ has $h_{min}(B_5^1) = 0.33 < 0.5 = h_{min}(G_5)$. This indicates that beam search may actually result in worse solutions than greedy search.

We proceed to show the instance where beam search obtains solutions with worse h_{sum} compared to greedy search in Example 2.

Example 2. Similar to Example 1, both algorithms start with one optimal pair, which is: $\{\pi_4, \pi_6\}$ with $h_{sum}(\{\pi_4, \pi_6\}) = 0.8$ and beam search starts with an addition pair $\{\pi_2, \pi_3\}$ with $h_{sum}(\{\pi_2, \pi_3\}) = 0.67$. In the next iteration, $B_3 = [\{\pi_3, \pi_4, \pi_6\}, \{\pi_1, \pi_2, \pi_3\}]$, $G_3 = \{\pi_3, \pi_4, \pi_6\}$, $h_{sum}(B_3^1) = h_{sum}(G_3) = 2.05$ and $h_{sum}(B_3^2) = 2.0167$. Then, in the next iteration, beam search commits to the $\{\pi_1, \pi_2, \pi_3\}$ branch as two of its successors outperform the top successor of $\{\pi_3, \pi_4, \pi_6\}$. As a result, $B_4 = [\{\pi_1, \pi_2, \pi_3, \pi_5\}, \{\pi_1, \pi_2, \pi_3, \pi_6\}]$, $G_4 = \{\pi_3, \pi_4, \pi_5, \pi_6\}$, and $h_{sum}(B_4^1) = 3.78$ and $h_{sum}(B_4^2) = 3.77 > h_{sum}(G_4) = 3.72$. Finally, we realized this was not a good decision as $B_5 = [\{\pi_1, \pi_2, \pi_3, \pi_4, \pi_5\}, \{\pi_1, \pi_3, \pi_4, \pi_5, \pi_6\}]$, $G_5 = \{\pi_1, \pi_3, \pi_4, \pi_5, \pi_6\}$, and $h_{sum}(B_5^2) = 5.87 < h_{sum}(B_5^1) = 5.90 < h_{sum}(G_5) = 5.97$.

C Additional Experiments on Bi-Criteria Framework

We present the omitted result from optimizing h_{sum} vs. optimizing a bi-criteria objective $f(S) = \alpha g_{sum}(S) + \beta h_{sum}(S)$ for $\alpha/\beta = 1/3$ (Figure 1) and $\alpha/\beta = 1$ (Figure 2). Figure 1 and Figure 2 show similar trends to what we showed in Section 5.2; however, the differences are less pronounced. The initial comparison within Figure 4 (left) reveals a slight decline in diversity under the bi-criteria framework, with the difference averaging around 1.76% (an average h_{sum} of 5.005 for directly optimizing h_{sum}, compared to an average h_{sum} of 4.917 achieved under the bi-criteria approach). In fact, directly optimizing for h_{sum} only led to superior diversity in 335 of the 1025 instances, with 643 instances recording equivalent h_{sum} values. Moving to the central graph in Figure 2, the bi-criteria framework was shown to be effective in enhancing g_{sum} of the solutions, though with fewer instances have benefited compared to the h_{min} and g_{min} setup. In particular, 327 instances observed improvements in g_{sum}, while 697 instances observed identical g_{sum}. Notably, in 502 of these ties, the selection of an optimal quality set was already achieved through h_{sum} optimization alone. The bi-criteria framework achieves an average g_{sum} of 4.097, outperforming the average g_{sum} of the diversity-only optimization (3.363) by about 21.82%. Finally, Figure 2 (right) demonstrates the strategic advantage of the bi-criteria approach, marking a noticeable improvement in the combined $g_{sum} + 3h_{sum}$ metric. Here, the bi-criteria optimization results in a 2.56%

average increase in the combined metric (18.848 versus 18.378 for optimization on pure diversity).

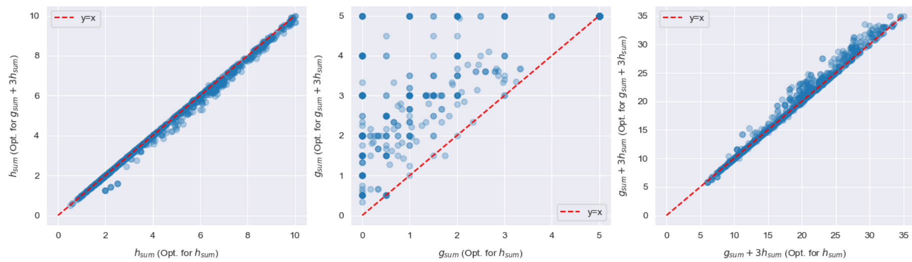

Fig. 1: Plots demonstrating the beam search (m = 128) performance of optimizing for h_{sum} vs. optimizing for $\alpha g_{sum} + \beta h_{sum}$ with $\frac{\alpha}{\beta} = \frac{1}{3}$.

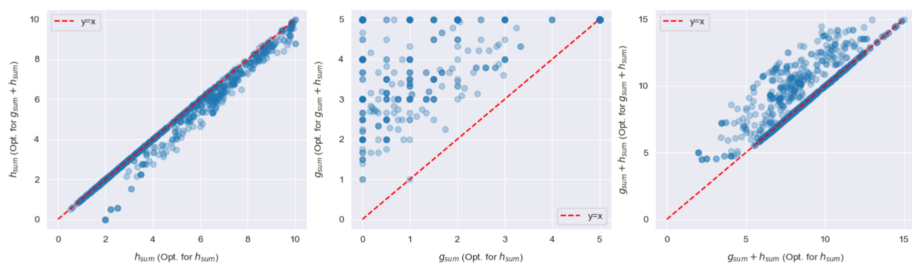

Fig. 2: Plots demonstrating the beam search (m = 128) performance of optimizing for h_{sum} vs. optimizing for $\alpha g_{sum} + \beta h_{sum}$ with $\frac{\alpha}{\beta} = 1$.

Figure 2 shows an alternative weighting scheme where the diversity and quality are equally weighted. As compared to Figure 1 (Left), we note that Figure 2 (Left) shows a larger decrease in diversity; in fact, the average h_{sum} is 4.783, which is about 5% lower than optimizing for diversity directly, which achieved an average h_{sum} of 5.005. In Figure 2 (Center), we see better gains in quality when not setting a diversity preference in the bi-criteria objective. In particular, using an equally weighted bi-criteria objective further improved the average g_{sum} to 4.487 vs. 4.097 achieved with a diversity-favored bi-criteria optimization vs. 3.363 achieved by directly optimizing for h_{sum}. As a result, we also see a much higher gain on the combined metric as less weight is put on diversity (9.270 vs 8.368, which roughly corresponds to an 11% improvement).

D MIP models

In this work, mixed integer programming (MIP) is only utilized to acquire provable optimal solutions as a reference, so its implementation is thus not heavily optimized. As a result, all MIP models from this work suffer, to some extent, from scalability issues, which remain for future work to explore.

The MIP model for the optimal $f(S) = h_{min}(S)$, the *minimum pairwise distance*, was originally proposed by [4]. Mathematically, the model is defined as:

$$\begin{aligned}
&\text{maximize } \{v_d : v_i \in \{0,1\}, i \in \{1, ..., n\}\} \\
&\text{subject to } \sum_{i=1}^{n} v_i = k, \\
&\qquad\qquad v_i + v_j + v_d \leq D_{ij} + 2, i \neq j
\end{aligned}$$

The model variables are defined as below:

- Binary variables v_i for every plan $\pi_i \in \Pi$. This serves as an indicator of whether a plan π_i is selected.
- A continuous variable v_d for bounding the pairwise diversity.

The constraints are defined as below:

1. Cardinality constraint $\sum_{\pi \in P} v_\pi = k$. This ensures that exactly k plans are being selected.
2. Pairwise constraint $\forall i < j \leq n : v_d + v_i + v_j \leq 2 + D_{ij}$ where D is the diversity matrix and D_{ij} is the pairwise diversity of plans $\pi_i, \pi_j \in \Pi$.

The objective of the model is then to maximize v_d. Unfortunately, due to the formulation of this model, we have at least $\binom{n}{2}$ pairwise constraints for a candidate set of n plans. This may lead to increased usage of time and memory, which will become an issue for large n.

We also propose a MIP model for $f(S) = h_{sum}(S)$, the *sum pairwise distance*. Mathematically, the model can be defined as:

$$\begin{aligned}
&\text{maximize } \{\sum_{i=1}^{n} \sum_{j \neq i}^{n} D_{ij} v_i v_j / 2 : v_i \in \{0,1\}, i \in \{1, ..., n\}\} \\
&\text{subject to } \sum_{i=1}^{n} v_i = k.
\end{aligned}$$

However, this model is a mixed integer quadratic program (MIQP) instead of a mixed integer linear program (MILP); unfortunately, we have not had the opportunity to optimize this model exclusively, as this is not the main focus of this work. The model variables are defined as below:

– Binary variables v_π for every plan $\pi \in P$. This serves as an indicator of whether a plan π is selected.

The constraints are defined as below:

1. Cardinality constraint $\sum_{\pi \in P} v_\pi = p$. This ensures that exactly p plans are being selected.

The objective function of the model is to maximize $v^T Dv$, where v is the vector representing plan selection, and D is the dispersion matrix such that $D_{ij} = d(\pi_i, \pi_j)$. This is equivalent to maximizing $\sum_{i,j} D_{ij} v_i v_j / 2$. Do note that We can linearize this model if we define additional $\binom{n}{2}$ binary variables z_{ij} and logical AND constraints to have $z_{ij} = 1$ if $v_i = v_j = 1$, but the model suffers from significant scalability issues as there are too many constraints.

Note all the above models do not contain quality metrics. To incorporate quality into the model, we can perform actions similar to how we incorporated diversity. In particular, for g_{min}, it is possible to add 1 auxiliary continuous variable v_c for bounding scaled plan cost and n linear constraints where $v_i + v_c \leq C_i + 1$, where C is the scaled cost vector. For g_{sum}, we can see that $v \cdot C$ precisely captures the metric. It should be noted that bringing cost into MIP can incur additional constraints and variables; this may worsen the scalability issue if not addressed.

References

1. Ravi, S.S., Rosenkrantz, D.J., Tayi, G.K.: Heuristic and special case algorithms for dispersion problems. Operations research **42**(2), 299–310 (1994)
2. Dasgupta, A., Kumar, R., Ravi, S.: Summarization through submodularity and dispersion. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 1014–1022 (2013)
3. Borodin, A., Lee, H.C., Ye, Y.: Max-sum diversification, monotone submodular functions and dynamic updates. In: Proceedings of the 31st ACM SIGMOD-SIGACT-SIGAI symposium on Principles of Database Systems. pp. 155–166 (2012)
4. Katz, M., Sohrabi, S., Udrea, O.: Bounding quality in diverse planning. In: Proceedings of the AAAI Conference on Artificial Intelligence. pp. 9805–9812 (2022)