Obtaining Diverse Solutions of Combinatorial Optimization Problems

See Woo Lee
Supervisor: Professor Akihiro Yamamoto

January 8th, 2021

1 Summary

The aim of this research is to investigate r-diverse combinatorial algorithms, which, given a set of feasible solutions $\mathcal{S}$ over a finite set $\mathcal{U}$, finds r feasible solutions maximizing a diversity measure defined by element-wise hamming distance. Thus far, the focus of the research has been r-diverse spanning trees and r-diverse bipartite matching. Both of these algorithms have polynomial time complexity regardless the value of r, which indicates the number of solutions.

2 Motivation

Algorithms that output *multiple* solutions instead of a single optimal solution enjoy relatively less attention, but they can be very useful when solving real-life problems for two important reasons.

First, often times the optimal solution is inapplicable for practical use, because many real-world constraints are simplified or even completely ignored in the mathematical model that a single-solution algorithm is designed optimize. A good example of this is finding the shortest route from one location to the other. Although an optimal pathfinding algorithm would indeed output the shortest possible path, it might be a bad idea to use this path in real life if it is affected by construction or suffers from heavier traffic compare to other, "less optimal" paths. Therefore, it is more practical to find multiple solutions that are sufficiently good rather than just one solution.

Second, many real life problem demands not just one solution but multiple good solutions. The best example of this is search engine, which returns thousands of relevant information given a search query [1].

One way to generate multiple good solutions is to use k-best enumeration algorithms, which produces a solution set of distinct k solutions that are more optimal than the other outputs not in the solution set [2]. However, if the k solutions are very similar to one another, the merit of the "k-best" approach diminishes. (To extend the example of pathfinding, all k paths generated by the algorithm may contain the road with heavy traffic congestion.)

For this reason, the diverse solution algorithms aim not only to find multiple good solutions, but also multiple solutions that are sufficiently different from one another [1].

3 Algorithms for Diverse Solutions

3.1 Diversity Measure

The diversity measure d mathematically quantifies how different solutions are from one another. In the works of Hanaka et al.[4] and Baste et al.[1], d is defined as the sum of pairwise Hamming distances among r solutions $U_1, \ldots, U_r \subseteq \mathcal{U}$ over a finite set $\mathcal{U}$.

$$d(U_1, \ldots, U_r) = \sum_{1 \leq i < j \leq r} (|U_i \setminus U_j| + |U_j \setminus U_i|)$$

3.2 Diverse Spanning Trees

Let $G = (V, E)$ be a graph and $E(G)$ be the edge set E of G. The objective of the r-diverse spanning tree algorithm is to find r spanning trees $T_1, \ldots, T_r$ of G such that $d(E(T_1), \ldots, E(T_r))$ is maximized. In this case, maximizing the pairwise Hamming distance signifies maximizing the symmetric difference between $E(T_i)$ and $E(T_j)$, which is equivalent to minimizing their intersection. This is implied by the following equation.

$$\begin{aligned}\mathcal{D} &= \sum_{1\le i<j\le r} (|E(T_i) \setminus E(T_j)| + |E(T_j) \setminus E(T_i)|) \\ &= \sum_{1\le i<j\le r} (|E(T_i)| + |E(T_j)| - 2|E(T_i) \cap E(T_j)|) \\ &= \sum_{1\le i<j\le r} 2(|V| - 1 - |E(T_i) \cap E(T_j)|)\end{aligned}$$

where $\mathcal{D} = d(U_1, \ldots, U_r)$. Exploiting the fact that $|V|$ is fixed and minimizing $|E(T_i) \cap E(T_j)|$ would maximize $\mathcal{D}$, the algorithm reduces this problem into finding r disjoint spanning trees with minimum weight given a weighted multigraph.

Following is the step-by-step procedure of the algorithm: First, given the graph G, we define a weighted multigraph $G' = (V, E')$. E' is created by duplicating each edge $e \in E$ with r parallel edges $e_1, \ldots, e_r$. Given $e \in E$, the duplicates $e_1, \ldots, e_r \in E'$ will be assigned weights of $0, \ldots, r-1$, respectively. Once such multigraph is constructed, Cunningham's algorithm is used to find the r disjoint weighted trees whose weights have an upper bound of k [4].

Creating the multigraph G' can be done in polynomial time, and so does running the Cunningham's algorithm. Overall, the time complexity of this algorithm is $\mathcal{O}((rn)^{2.5}m)$, where n represents the number of vertices and m, the number of edges in G [4].

3.3 Diverse Bipartite Matching

Let $G = (A \cup B, E)$ be a bipartite graph where A and B are color classes of G. Let k and r be integers. Then r-diverse bipartite matching algorithm computes r matchings $M_1, \ldots, M_r$ of the graph G with the cardinality k that maximizes $d(M_1, \ldots, M_r)$. The algorithm computes the solutions by reducing the problem to the minimum cost flow problem.

Similar to the diverse spanning tree algorithm, we create $G' = (V', E')$, where G' is a weighted multigraph version of the original graph G. However, G' has few other modifications as well. First, $\{a, b\} \in E'$ are arcs from a to b rather than undirected edges, where $a \in A$ and $b \in B$. Second, a source vertex s and a sink vertex t is added to V to form V'; then the arcs from s to every $a \in A$ and from every $b \in B$ to t are created, with the weight of 0. With these additional changes made, the minimum cost flow algorithm can be utilized to derive the minimum weight subgraph H^*, which has the maximum degree of at most r and kr edges. H^* can be partitioned into r bipartite matchings $M_1, \ldots, M_r$ by ensuring that all of the matching has the same cardinality.

The research of Fomin et al. [3] has demonstrated that r-diverse bipartite matching has polynomial time complexity when $r = 2$. This version of the algorithm, however, enjoys polynomial time complexity regardless the output size r, just like the diverse spanning tree algorithm.

4 Future Research

The research on other polynomial-time tractable diverse algorithms, such as diverse shortest st-paths, will be continued. Given a graph $G = (V, E)$ and a starting point s and a destination t, the diverse shortest st-paths algorithm will output r shortest paths $P_1, \ldots, P_r$ from s to t that maximize $d(P_1, \ldots, P_r)$.

The scope of the research may be expanded to FPT diverse algorithms, where the fixed parameter is the number of diverse solutions.

References

[1] Julien Baste et al. "Diversity of Solutions: An Exploration Through the Lens of Fixed-Parameter Tractability Theory". In: *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence* (July 2020). DOI: 10.24963/ijcai.2020/156. URL: http://dx.doi.org/10.24963/ijcai.2020/156.

[2] David Eppstein. "K-Best Enumeration". In: *Bull. EATCS* 115 (2015). URL: http://eatcs.org/beatcs/index.php/beatcs/article/view/322.

[3] Fedor V. Fomin et al. "Diverse Pairs of Matchings". In: *31st International Symposium on Algorithms and Computation, ISAAC 2020, December 14-18, 2020, Hong Kong, China (Virtual Conference)*. Ed. by Yixin Cao, Siu-Wing Cheng, and Minming Li. Vol. 181. LIPIcs. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020, 26:1–26:12. DOI: 10.4230/LIPIcs.ISAAC.2020.26. URL: https://doi.org/10.4230/LIPIcs.ISAAC.2020.26.

[4] Tesshu Hanaka et al. *Finding Diverse Trees, Paths, and More*. 2020. arXiv: 2009.03687 [cs.DS].