THE ABSTRACT OF DOCTORAL THESIS

1. THESIS INFORMATION

- Full name: Tran Dinh Tan

- Thesis title: Development of Approximation Algorithms for Submodular Function Maximization.

- Major: Computer sciense Code: 9480101

- Education facility: VNU University of Engineering and Technology, Vietnam National University, Hanoi

2. CONTENT

The thesis focuses on the study of **approximation algorithms for submodular function optimization problems**. The core approach lies in exploiting the special mathematical structure of submodular functions, combined with approximation algorithm design techniques, to develop algorithms with rigorous theoretical guarantees in terms of approximation ratios and query complexity. Moreover, the thesis emphasizes the practical implementation on real-world datasets to evaluate the effectiveness of the proposed algorithms for large-scale problems. The overarching perspective is to establish methods that balance theoretical performance with practical applicability. Based on the theoretical gaps identified in submodular function optimization, the thesis investigates four representative problems:

1. Submodular Maximization under Knapsack Constraint: Given a ground set V, a submodular function $f: 2^V \to R^+$, and a cost function $c(e) > 0$ for each element $e \in V$, the cost of a subset $S \subseteq V$ is defined as $c(S) = \sum_{e \in S} c(e)$. Given a budget B, the objective is to find a subset $S \subseteq V$ such that $f(S)$ is maximized under the constraint $c(S) \leq B$. This thesis develops three algorithms - **DLA, RLA**, and **AST** - with approximation ratios of $1/6 - \epsilon, 1/4 - \epsilon \ and \ 1/7 - \epsilon$, respectively. The results have been published at two A*-ranked international conferences: **IJCAI 2023** and **IJCAI 2024**.

2. k-Submodular Maximization under Individual Knapsack Constraints: Given a ground set V, a k-submodular function $f: (k+1)^V \to R_+$ defined over the space of a k-set $\boldsymbol{s} = (S_1, S_2, \dots, S_k)$, where each $S_i \subseteq \mathrm{V}, S_i \cap S_j = \emptyset$ for all $i \neq j$, and each element $e \in V$ has a cost $c(e) > 0$. For each index $i \in [k] = \{1,2,\dots,k\}$, the total cost of the set S_i is defined as $c_i(\boldsymbol{s}) = \sum_{e \in S_i} c(e)$, and each index i has a individual budget $B_i > 0$. The objective is to find a k-tuple $\boldsymbol{s} = (S_1, S_2, \dots, S_k)$ such that $c_i(s) \leq B_i, \forall i \in [k]$, and the objective value $f(\boldsymbol{s})$ is maximized. This thesis proposes a streaming algorithm, **STR**, which achieves an approximation ratio of $\frac{1-\epsilon}{2(k+1)}$ in the monotone case and $\frac{1-\epsilon}{2k+3}$ in the non-monotone case, with low query and memory complexity. The algorithm is well-suited for large-scale data environments and represents the first efficient streaming approach for the k-submodular maximization problem under individual knapsack constraints. The research outcomes have been published in the international conference **SOICT 2023 (SCOPUS)** and the international journal **APJOR (SCIE, Q3)**.

3. DR-submodular Maximization under Size Constraint: Given a ground set V, a DR-submodular function $f: Z_+^E \mapsto R_+$, a bounded integer lattice $\mathbb{B}$, and a positive integer $k > 0$, the objective is to find a vector $\boldsymbol{x} \leq \mathbb{B}$ such that the total size $|\boldsymbol{x}|_1 \leq k$ and the function value $f(\boldsymbol{x})$ is maximized. This thesis develops two novel algorithms, **FastDrSub** and **FastDrSub+**, which achieve approximation ratios of 0.044 and $\frac{1}{4} - \epsilon$, respectively, with low query complexity. These are the first algorithms to attain constant-factor approximation for this problem, and they exhibit outstanding performance in the Revenue Maximization problem with DR-submodular objectives. The results have been published in the international journal **JOCO (SCIE, Q2).**

4. Minimum cost Submodular Cover: Given a ground set V, a monotone submodular function $f: 2^V \rightarrow R^+$, and a cost function $c(e) > 0$ for each element $e \in V$, the cost of a subset $S \subseteq V$ is defined as $c(S) = \sum_{e \in S} c(e)$. Given a quality threshold $T > 0$, the objective is to find a subset $S \subseteq V$ such that $f(S) \geq T$ and $c(S)$ is minimized. This thesis develops three efficient streaming algorithms—**SingStr**, **ThreeStr**, and **MultiStr**—that enable processing large-scale streaming data. These algorithms achieve approximation ratios close to the Greedy method while significantly reducing the number of queries and memory usage. Moreover, they demonstrate practical effectiveness in applications such as Revenue threshold and Coverage threshold. The results have been published at the international conference **CSONET 2023** (**SCOPUS**).

The four studied problems represent a diversity of objective function types and constraint types, thereby outlining a comprehensive landscape of the submodular optimization problem space. The selection of problems that both extend foundational results and explore novel problem settings aims to balance breadth and depth in research. From a theoretical standpoint, this thesis contributes several novel results with clear academic significance. Specifically, the proposed algorithms achieve strong approximation ratios and yield substantial improvements over existing methods. The thesis also provides a detailed analysis of parallel complexity and the number of oracle queries required, supporting the applicability of these algorithms to large-scale problems. These findings have been published in multiple reputable peer-reviewed international conferences and journals.