

# TRÍCH YẾU LUẬN ÁN

## 1. TÓM TẮT MỞ ĐẦU

- Họ và tên NCS: Trần Đình Tân
- Tên luận án: Phát triển thuật toán xấp xỉ cho bài toán tối đa hàm submodular
- Ngành: Khoa học máy tính Mã số: 9480101
- Đơn vị đào tạo: Trường Đại học Công nghệ, Đại học Quốc gia Hà Nội

## 2. NỘI DUNG

Luận án tập trung nghiên cứu chủ đề **các thuật toán xấp xỉ cho bài toán tối đa hàm submodular**. Hướng tiếp cận cốt lõi là khai thác cấu trúc toán học đặc biệt của hàm submodular và kết hợp với các phương pháp thiết kế thuật toán xấp xỉ để thiết kế các thuật toán có đảm bảo lý thuyết rõ ràng về hệ số xấp xỉ, độ phức tạp truy vấn. Đồng thời, luận án chú trọng đến khả năng triển khai thực nghiệm trên dữ liệu thực, nhằm đánh giá hiệu quả thuật toán trong các bài toán có quy mô lớn. Quan điểm xuyên suốt là xây dựng các phương pháp cân bằng giữa hiệu năng lý thuyết và khả năng ứng dụng trong thực tiễn. Dựa khoảng trống lý thuyết được phân tích cho bài toán tối ưu hàm submodular, luận án đã lựa chọn và nghiên cứu bốn bài toán tiêu biểu:

**1. Tối đa hàm submodular với ràng buộc chi phí:** Cho một tập cơ sở  $V$ , một hàm submodular  $f: 2^V \rightarrow R^+$  và mỗi phần tử  $e \in V$  có chi phí  $c(e) > 0$ . Chi phí của một tập con  $S \subseteq V$  là  $c(S) = \sum_{e \in S} c(e)$ . Với một ngưỡng chi phí  $B$ , mục tiêu là tìm một tập con  $S \subseteq V$  sao cho:  $f(S)$  lớn nhất và  $c(S) \leq B$ . Luận án đã phát triển ba thuật toán **DLA**, **RLA** và **AST** với hệ số xấp xỉ lần lượt là  $1/6 - \epsilon$ ,  $1/4 - \epsilon$  và  $1/7 - \epsilon$ . Các thuật toán này không chỉ đạt độ phức tạp truy vấn tuyến tính hoặc song song tối ưu, mà còn thể hiện hiệu quả vượt trội trong các môi trường thực nghiệm. Các kết quả đã được công bố tại hai hội nghị quốc tế xếp hạng A\* là **IJCAI 2023** và **IJCAI 2024**.

**2. Tối đa hàm  $k$ -submodular với ràng buộc chi phí nhóm:** Cho một tập cơ sở  $V$ , một hàm  $k$ -submodular  $f: (k+1)^V \rightarrow R_+$  xác định trên không gian của các  $k$ -tập  $\mathbf{s} = (S_1, S_2, \dots, S_k)$  với  $S_i \subseteq V, S_i \cap S_j = \emptyset$  nếu  $i \neq j$  và mỗi phần tử  $e \in V$  có chi phí  $c(e) > 0$ . Với mỗi vị trí  $i \in [k] = \{1, 2, \dots, k\}$ , tổng chi phí của tập  $S_i$  được định nghĩa là  $c_i(\mathbf{s}) = \sum_{e \in S_i} c(e)$ , và mỗi vị trí có một ngân sách riêng  $B_i > 0$ . Bài toán yêu cầu tìm một  $k$ -tập  $\mathbf{s} = (S_1, S_2, \dots, S_k)$  sao cho:  $c_i(\mathbf{s}) \leq B_i, \forall i \in [k]$ , và giá trị hàm mục tiêu  $f(\mathbf{s})$  được tối đa. Luận án đề xuất thuật toán dòng **STR**, đạt hệ số xấp xỉ  $\frac{1-\epsilon}{2(k+1)}$  trong trường hợp đơn điệu và  $\frac{1-\epsilon}{2k+3}$  trong trường hợp không đơn điệu, với độ phức tạp truy vấn và bộ nhớ thấp. Thuật toán này phù hợp với môi trường dữ liệu lớn và là hướng tiếp cận đầu tiên hiệu quả cho bài toán tối ưu  $k$ -submodular trong bối cảnh ràng buộc chi phí nhóm. Kết quả nghiên cứu đã được công bố trên hội thảo quốc tế **SOICT 2023(SCOPUS)** và tạp chí quốc tế **APJOR (SCIE, Q3)**.

**3. Tối đa hàm DR-submodular với ràng buộc lực lượng:** Cho một tập cơ sở  $V$ , một hàm DR-submodular  $f: Z_+^E \mapsto R_+$ , một lưới nguyên bị chặn bởi vector  $\mathbb{B}$ , và một số nguyên dương  $k > 0$ . Bài toán yêu cầu tìm vector  $\mathbf{x} \leq \mathbb{B}$  sao cho tổng kích thước  $|\mathbf{x}|_1 \leq k$  và giá trị  $f(\mathbf{x})$  đạt cực đại. Luận án phát triển hai thuật toán mới là **FastDrSub** và **FastDrSub+**, lần lượt đạt hệ số xấp xỉ 0.044 và  $\frac{1}{4} - \epsilon$  với độ phức tạp truy vấn thấp. Đây là những thuật

toán đầu tiên đạt được xấp xỉ hằng số cho bài toán này, đồng thời thể hiện hiệu quả vượt trội trong bài toán tối đa doanh thu với hàm mục tiêu DR-submodular. Kết quả đã được chấp nhận công bố tại tạp chí **JOCO(SCIE, Q2)**.

**4. Bài toán phủ submodular với chi phí tối thiểu** (đối ngẫu của bài toán 1): Cho một tập cơ sở  $V$ , một hàm submodular đơn điệu  $f: 2^V \rightarrow R^+$  và mỗi phần tử  $e \in V$  có chi phí  $c(e) > 0$ . Chi phí của một tập con  $S \subseteq V$  là  $c(S) = \sum_{e \in S} c(e)$ . Với một ngưỡng chất lượng  $T > 0$ , mục tiêu là tìm một tập con  $S \subseteq V$  sao cho:  $f(S) \geq T$  và  $c(S)$  là nhỏ nhất. Luận án đã xây dựng ba thuật toán dòng hiệu quả là **SingStr**, **ThreeStr** và **MultiStr**, cho phép xử lý dữ liệu lớn theo dòng. Các thuật toán đạt hệ số xấp xỉ gần bằng tham lam nhưng sử dụng ít truy vấn và bộ nhớ hơn đáng kể, đồng thời thể hiện tính hiệu quả trong các ứng dụng như ngưỡng doanh thu và ngưỡng phủ. Các kết quả đã được công bố tại hội nghị quốc tế **CSOINET 2023 (SCOPUS)**.

Bốn bài toán nghiên cứu đại diện cho sự đa dạng về loại hàm mục tiêu, tính đơn điệu và dạng ràng buộc, từ đó hình thành một bức tranh tổng thể về không gian bài toán tối ưu submodular. Việc lựa chọn cả bài toán vừa bao phủ các trường hợp đã có kết quả nền để cải tiến, vừa khai phá các bài toán chưa từng được nghiên cứu, đảm bảo tính cân đối giữa chiều sâu và chiều rộng nghiên cứu. Về mặt lý thuyết, luận án đã đóng góp nhiều kết quả mới có giá trị học thuật rõ ràng. Cụ thể, các thuật toán được đề xuất đều đạt hệ số xấp xỉ tốt và cải thiện đáng kể so với các phương pháp hiện có. Luận án cũng đưa ra các phân tích chi tiết về độ phức tạp song song và số lượng truy vấn cần thiết, hỗ trợ việc áp dụng cho các bài toán lớn. Các kết quả này đã được công bố tại nhiều hội nghị và tạp chí quốc tế uy tín có phản biện.