1. **Trình bày Fuzzy C-Mean**

Thông thường, mỗi phương pháp phân cụm dữ liệu phân một tập dữ liệu ban đầu thành các cụm dữ liệu có tính tự nhiên và mỗi đối tượng dữ liệu chỉ thuộc về một cụm dữ liệu. Phương pháp này chỉ phù hợp với viêc khám phá ra các cụm có mật độ cao và rời nhau.

Tuy nhiên, trong thực tế các cụm dữ liệu lại có thể chồng lên nhau, nghĩa là một số các đối tượng dữ liệu thuộc về nhiều cụm khác nhau, người ta áp dụng lý thuyết tập mờ trong phân cụm dữ liệu để giải quyết trường hợp này, cách thức kết hợp này gọi là phân cụm mờ.

Thuật toán phân cụm mờ C-Means được kế thừa và phát triển tư tưởng của thuật toán phân cụm rõ K-Means. Cả hai thuật toán này đều sử dụng chung một chiến lược đó là phân cụm dữ liệu. Thuật toán phân cụm dữ liệu mờ C-means hay còn gọi tắt là thuật toán Fuzzy C-Means đã được áp dụng thành công trong việc giải quyết một số bài toán như trong nhận dạng mẫu, xử lý ảnh, y học, thống kê, … .

Diagram

Description automatically generatedLưu đồ của thuật toán Fuzzy C Means

1. **Code Fuzzy C Mean và demo trên dữ liệu mẫu**

Import thư viện



Tạo dữ liệu giả

Chart, scatter chart

Description automatically generated

Xây dựng mô hình Fuzzy C-Means



Xem kết quả



Chart, scatter chart

Description automatically generated

Hình bên trái thể hiện dữ liệu gốc, hình bên phải mô tả 2 cụm của dữ liệu sau khi sử dụng Fuzzy C-Means

1. **Hướng giải quyết bài toán tư vấn chọn chuyên ngành.**

Chương trình đào tạo hệ đại học chính quy ngành Công nghệ thông tin được chia làm 4 chuyên ngành gồm: Hệ thống thông tin, Công nghệ phần mềm, Mạng máy tính và Khoa học dữ liệu. Mỗi chuyên ngành gồm những học phần chuyên sâu thể hiện khối kiến thức đặc thù của chuyên ngành đó. Hàng năm, Khoa Công nghệ Thông tin thường tổ chức buổi giới thiệu chuyên ngành cho sinh viên năm thứ 3. Qua đó, sinh viên sẽ xác định được chuyên ngành nào phù hợp và tiến hành đăng ký môn học theo chuyên ngành đã chọn. Tuy nhiên, có thể thấy việc lựa chọn chuyên ngành như trên phần lớn là cảm tính, theo sở thích của sinh viên mà chưa có căn cứ cụ thể dẫn đến việc chọn chuyên ngành không phù hợp gây ảnh hưởng lớn đến kết quả học tập của sinh viên. Nhóm tác giả đưa ra giải pháp và thực nghiệm giải quyết việc định hướng chuyên ngành cho sinh viên một cách tự động căn cứ vào kết quả học tập những học phần có kiến thức bổ trợ cho từng chuyên ngành. Các học phần này được chọn theo từng chuyên ngành như sau:

* **Khoa học dữ liệu**: Nhập môn lập trình, toán rời rạc, khai thác dữ liệu, cơ sở dữ liệu
* **Hệ thống thông tin**: Cơ sở dữ liệu, hệ quản trị cơ sở dữ liệu, nhập môn lập trình, cấu trúc dữ liệu
* **Công nghệ phần mềm**: Nhập môn lập trình, lập trình hướng đối tượng, cơ sở dữ liệu
* **Mạng máy tính**: Kiến trúc máy tính, mạng máy tính, nhập môn lập trình, hệ điều hành

Ứng dụng Fuzzy C-Means trong phân cụm chuyên ngành

Thuật toán Fuzzy C-Means được áp dụng phân cụm cho từng chuyên ngành, mỗi chuyên ngành có kết quả là 2 cụm, một cụm gồm các sinh viên có khả năng theo học chuyên ngành đó và cụm còn lại là sinh viên không có khả năng học, lặp lại thuật toán cho đến hết các chuyên ngành. Trọng tâm ban đầu của mỗi cụm trong từng chuyên ngành được chỉ định với 2 mức gọi là ngưỡng trên và ngưỡng dưới. Ngưỡng trên sẽ gom cụm gồm những sinh viên có khả năng học chuyên ngành, ngưỡng dưới gom cụm gồm những sinh viên không có khả năng học chuyên ngành.

Thuật toán Fuzzy C-Means áp dụng vào bài toán được trình bày như sau:

Input:

* Bảng điểm sinh viên tổng hợp đã qua bước tiền xử lý
* Danh sách các môn học được chọn theo từng chuyên ngành.
* Trọng tâm của 2 cụm ứng với ngưỡng trên (có khả năng) và ngưỡng dưới (không có khả năng)

Output: Danh sách sinh viên được phân cụm theo từng chuyên ngành và danh sách sinh viên không thuộc chuyên ngành nào.

Bước 1: Khởi tạo

* Trọng tâm ban đầu theo từng chuyên ngành trong đó ki là trọng tâm của cụm thuộc chuyên ngành ; là điểm số thứ thuộc trọng tâm .
* Danh sách điểm số trong đó ki là nhóm điểm thứ thuộc chuyên ngành ; là điểm số thứ thuộc nhóm điểm .

Bước 2: Phân cụm cho mỗi chuyên ngành

For k = 1 to N do //Lặp N chuyên ngành

= Fuzzy C-Means (); //Xử lý phân cụm cho chuyên ngành

// gồm 2 tập: gồm những sinh viên được phân cụm vào chuyên ngành , gồm những sinh viên không thuộc chuyên ngành

Bước 3: Xử lý kết quả

For i = 1 to M do //Duyệt qua danh sách M sinh viên đầu vào

Nếu ∉ {tập tất cả các chuyên ngành đã phân cụm}

∈ {Danh sách khác}