

Đề xuất đề tài:

Gợi ý phim

1. Danh sách thành viên nhóm

Họ tên	MSSV	Email
Nguyễn Ngọc Đôn	20130941	ngocdon0127@gmail.com
Nguyễn Tất Hòa	20131536	kevinhoa95@gmail.com
Bùi Văn Toàn	20134033	toanshikoda@gmail.com
Đinh Thế Anh	20130046	anhdtav95@gmail.com

2. Đề xuất đề tài

- Mô tả bài toán thực tế

Với một tập các phim mà một người dùng đã xem và đánh giá, hệ thống cần phải xác định (dự đoán) những bộ phim nào (chưa được xem) mà người dùng đó thích xem.

- Ý tưởng

Hai người xem những bộ phim giống nhau, đánh giá các phim đó giống nhau thì sẽ có xu hướng thích cùng các bộ phim khác trong tương lai; người này sẽ thích những bộ phim mà người kia đã đánh giá cao và ngược lại.

- Cách thức biểu diễn dữ liệu

Dữ liệu học được biểu diễn bằng các ví dụ học có cấu trúc: (u_i, m_i, r_i) với r_i là giá trị (số thực từ 0.5 đến 5, các mức đánh giá cách nhau 0.5) mà user u_i đánh giá cho bộ phim m_i .

- **Các giải thuật học máy được sử dụng**

Giải thuật học máy được sử dụng là giải thuật học dựa trên láng giềng gần nhất – K-nearest neighbors (k-NN).

Với mỗi cặp (u, m) cần dự đoán, ta sẽ tính toán K người dùng tương tự nhất đối với người dùng u , và tính giá trị đánh giá trung bình của K người dùng đó đối với phim m để dự đoán đánh giá của người dùng u đối với phim.

- **Input – Output**

Input: Nhập vào danh sách các bộ phim và đánh giá của 1 người dùng.

Output: Xuất ra danh sách những bộ phim mà người dùng đó có thể sẽ thích xem.

- **Tập dữ liệu được sử dụng**

<http://grouplens.org/datasets/movielens/20m/>

Có thể sử dụng thêm 1 phần dữ liệu của Netflix Prize